Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Écologie et Évolution*

Présentée et soutenue par *LISA FOURTUNE*
Le *12 janvier 2018*

Titre :

Patrons de diversité inter- et intraspécifique dans les réseaux dendritiques d'eau douce : implications pour leur fonctionnement et leur conservation

---

JURY

Sébastien Brosse (président du jury)
Florian Altermatt (rapporteur)
Christine Argillier (rapporteur)
Guillaume Evanno (examinateur)
Gaël Grenouillet (examinateur)
Simon Blanchet (directeur de thèse)

Ecole doctorale : *SEVAB*
Unité de recherche : *UMR 5321 - Station d'Écologie Théorique et Expérimentale*
Directeur(s) de Thèse : *Simon Blanchet*
Rapporteurs : *Florian Altermatt et Christine Argillier*

Résumé

L'ensemble des différentes facettes de la biodiversité connaissent actuellement un fort déclin dû à l'action de l'Homme. Dans ce contexte, le principal défi qui se pose aux scientifiques est de proposer aux décisionnaires des solutions efficaces et durables pour limiter les pertes de biodiversité. Faire face à ce défi nécessite la pleine compréhension des nombreuses facettes de la biodiversité (biodiversité interspécifique, intraspécifique, des écosystèmes), ainsi que de ses différents niveaux (diversité aux niveaux α, β et γ). Notamment, il est nécessaire de (i) savoir comment la biodiversité est distribuée spatialement, (ii) mettre en évidence quels processus évolutifs et écologiques sont à l'origine de cette distribution, et (iii) isoler les possibles covariations spatiales et interactions entre les différentes facettes de la biodiversité.

Décrire la distribution spatiale de la biodiversité et identifier les processus qui la sous-tendent est un défi statistique majeur, car cela implique d'isoler –*in situ*- des relations causales complexes entre de nombreuses variables. Une solution possible est l'utilisation de modèles causaux. Les méthodes actuelles (telles que les méthodes de *path analysis* et le *d-sep test*) semblent en effet tout à fait adaptées à l'étude des patrons de biodiversité au niveau α. Cependant, au niveau β, les variables prennent la forme de matrices de distances (par exemple, les matrices de différenciation génétique ou taxonomique entre sites) et les méthodes actuelles ne peuvent être appliquées. Le premier chapitre de cette thèse a donc été consacré au développement de nouvelles méthodes statistiques permettant l'analyse, par des modèles causaux, de données sous la forme de matrices de distances.

Dans le deuxième chapitre de cette thèse, j'ai étudié les covariations spatiales entre diversité interspécifique et diversité génétique intraspécifique chez quatre espèces de poissons de rivières (*Barbatula barbatula*, *Gobio occitaniae*, *Phoxinus phoxinus* et *Squalius cephalus*). Au niveau α, nous avons mis en évidence des corrélations positives entre diversité interspécifique et diversité génétique intraspécifique chez les quatre espèces, et, au niveau β, nous avons constaté des corrélations plus faibles, et qui n'étaient significatives que chez deux des quatre espèces. L'utilisation de modèles causaux nous a permis de démontrer que des processus évolutifs et écologiques communs (tel que le filtrage environnemental, la migration et la dérive) affectaient les diversités interspécifique et intraspécifique au sein des communautés de poissons de rivière.

Dans le troisième chapitre de cette thèse, j'ai étudié les corrélations entre diversité génétique (neutre) et phénotypique à l'échelle intraspécifique chez deux espèces de poissons de rivière (*G. occitaniae* et *P. phoxinus*). Au niveau α, les corrélations entre diversité génétique et phénotypique étaient non-significatives, et au niveau β, nous avons constaté une corrélation positive chez une des deux espèces seulement. Comme attendu, la diversité génétique était principalement déterminée par des processus neutres alors que la diversité phénotypique était liée à des processus adaptatifs. Néanmoins, par des analyses causales, nous avons mis en évidence que la nature des processus impliqués différait entre les espèces.

Dans le quatrième et dernier chapitre de cette thèse, j'ai utilisé un modèle de dynamique éco-évolutive afin de mettre en évidence l'impact des principales caractéristiques des réseaux dendritiques d'eau douce (la structure dendritique, les gradients de capacité de charge et de conditions environnementales entre l'amont et l'aval, et la migration asymétrique dûe au courant) sur l'adaptation locale et la distribution de la diversité phénotypique. Nous avons constaté que, alors que la structure dendritique en elle-même ne semblait pas influer fortement sur l'adaptation locale, les gradients de capacité de charge et de conditions environnementales semblaient avoir un impact fort sur les patrons d'adaptation, notamment lorsqu'ils étaient combinés.

Abstract

All facets of biodiversity are currently facing a dramatic decline due to human activities. In this context, a main challenge faced by scientists is to propose to decision-makers efficient and sustainable plans for limiting biodiversity loss. This challenge requires an extensive understanding of the many facets (i.e. intraspecific diversity, interspecific diversity and diversity of ecosystems) and components (i.e. components α, β and γ) of biodiversity. Most notably, further knowledge are required regarding (i) how biodiversity is spatially distributed, (ii) what are the ecological and evolutionary processes shaping the spatial distribution of biodiversity and (iii) how the different facets of biodiversity are interacting with one another.

Describing the spatial distribution of biodiversity and understanding its underlying drivers represent major statistical challenges as it implies disentangling -in the wild- intricate causal relationships among numerous factors. A solution may build on methods of causal modeling. The existing methods of causal modeling are well suited to study biodiversity patterns at the α-level. However, at the β-level, the variables take the form of pairwise matrices (e.g. matrix of genetic or taxonomic differentiation among sites) and these methods cannot be used. The first chapter of this thesis aimed at developing novel statistical approaches allowing the application of two methods of causal modeling (namely path analysis and the d-sep test) to data taking the form of pairwise matrices.

In the second chapter of this thesis, I studied the spatial covariation between interspecific diversity and intraspecific neutral genetic diversity patterns (named Species-Genetic Diversity Correlations, SGDCs) in four freshwater fish species (*Barbatula barbatula*, *Gobio occitaniae*, *Phoxinus phoxinus* and *Squalius cephalus*). I found significant and moderate positive SGDCs at the α-level for all four fish species, whereas at the β-level, SGDCs were weaker in strength and positively significant for two out of the four species. Using causal modeling, I showed that similar evolutionary and ecological processes related to environmental filtering, migration, drift and colonization history shaped both the interspecific and intraspecific diversity of fish communities.

In the third chapter of this thesis, I studied the correlations between (neutral) genetic and phenotypic diversity at the intraspecific level (named Genetic-Phenotypic Intraspecific Diversity Correlations, GPIDCs) in two freshwater fish species (*G. occitaniae* and *P. phoxinus*). We found no GPIDCs at the α-level and a positive GPIDC at the β-level in *G. occitaniae* only. As expected, genetic diversity was mainly driven by neutral processes and phenotypic diversity was driven by adaptive processes, although the nature of these processes differed between species. At the β-level, the positive GPIDC appeared to originate from a direct relationship between the two levels of biodiversity.

In the fourth and last chapter of this thesis, I used an eco-evolutionary metapopulation dynamics model to assess the impacts of the main features of riverine networks (i.e. the dendritic structure, the upstream-downstream gradient in habitat capacities, the upstream-downstream gradient in environmental conditions and the asymmetric dispersal rate caused by water flow) on local adaptation and the distribution of intraspecific diversity. I found that, although the dendritic structure in itself did not strongly influence local adaptation, gradients in habitat capacities and environmental conditions had an important impact on local adaptation, and that this impact was even stronger when they were combined.

Avant-Propos

Cette thèse a été financée par le Ministère de l'Enseignement Supérieur et de la Recherche, et par la Station d'Ecologie Théorique et Expérimentale (CNRS - Université Toulouse 3).

Ce document de thèse débute par une introduction générale rédigée en anglais ; elle est suivie de quatre chapitres présentés sous forme d'articles scientifiques en anglais, puis d'une conclusion générale en anglais.

L'introduction générale a pour objectif de situer la thèse dans le contexte actuel de la recherche en écologie et évolution.

Les quatre chapitres suivants font ou vont faire l'objet d'une publication dans une revue scientifique à comité de lecture.

Le chapitre I a été accepté pour publication après révisions mineures dans The American Naturalist.

Le chapitre II a été publié dans la revue Freshwater Biology en 2016.

Le chapitre III est en préparation pour soumission dans la revue Journal of Animal Ecology. La présente version n'a pas encore été présentée à la relecture de tous les co-auteurs.

Le chapitre IV est en préparation pour soumission dans une revue à comité de lecture. La présente version n'a pas encore été présentée à la relecture de tous les co-auteurs.

Simon Blanchet est le directeur de cette thèse, et co-auteur de tous les articles présentés. Il a contribué au développement pratique, théorique et analytique de ces chapitres, à la récolte des données et à la rédaction des articles scientifiques.

Géraldine Loot est co-directrice de cette thèse. Elle a participé au développement et à la rédaction des chapitres I à III, dont elle est co-auteur.

Jérome G. Prunier est co-auteur de tous les chapitres de cette thèse. Il a contribué au développement théorique de tous les chapitres, ainsi qu'à l'analyse des données génétiques utilisées dans les chapitres II et III.

Ivan Paz-Vinas a participé à la rédaction des chapitres I et II dont il est co-auteur.

Charlotte Veyssière est co-auteure des chapitres I et III. Elle a assuré la grande majorité des analyses moléculaires nécessaire à l'élaboration des bases de données génétiques utilisées dans ces articles.

Nicolas Canto est co-auteur du chapitre III. Il a grandement contribué à la mise en place et la réalisation de la collecte d'échantillons sur le terrain ayant permis la création de la base de données sur laquelle se base ce chapitre, ainsi qu'à l'extraction des données morphologiques.

Eglantine Mathieu-Bégné est co-auteure du chapitre III. Elle a contribué à l'extraction des données morphologiques sur lesquelles se base ce chapitre.

Claire de Mazancourt et Bart Haegeman sont co-auteurs du chapitre IV. Ils ont participé au développement théorique de cet article.

De nombreuses personnes ayant fait ou faisant partie de l'équipe EEDYL de la Station d'Ecologie Théorique et Expérimentale (CNRS - Université Toulouse 3), du laboratoire Evolution et Diversité Biologique (Université Toulouse 3), ainsi que plusieurs équipes de l'ONEMA, ont participé à la collecte d'échantillons sur le terrain, collaborant ainsi à la création des chapitres I, II et III.

Remerciements

Tout d'abord, je tiens à remercier Florian Altermatt et Christine Argillier pour avoir accepté de consacrer une partie de leur temps à la lecture de ce manuscrit de thèse ainsi qu'à l'évaluation de mon travail. Je remercie également Sébastien Brosse, Guillaume Evanno et Gaël Grenouillet qui ont accepté de faire partie de mon jury de soutenance.

Je remercie du fond du cœur Simon Blanchet et Géraldine Loot pour leur disponibilité, leur confiance et leur soutien, même dans les coups durs. Merci à Jérome G. Prunier pour son aide et ses conseils avisés. Merci à Ivan Paz-Vinas pour son soutien. Merci à Claire de Mazancourt et Bart Haegeman pour leur apport précieux.

Ma gratitude va également à tous les gens ayant rendu possibles la collecte et le traitement des données sur lesquelles se base le présent travail. Merci notamment à Nicolas Canto pour le travail de terrain et l'analyse des données morphologiques, Charlotte Veyssière pour les analyses génétiques et Eglantine Mathieu-Bégné pour l'extraction d'une partie des données morphologiques.

Enfin je tiens à remercier toutes les personnes qui ont participé au travail de terrain, et notamment Olivier Rey, Elise Mazé et Kéoni Saint-Pé ainsi que les équipes de l'ONEMA avec lesquelles nous avons collaboré.

Enfin, mes remerciements vont à Vincent Dubut, Sébastien Villéger et Marie-Hélène Lizée pour l'aide et les conseils qu'ils m'ont apportés lors de mes comités de suivi de thèse.

Je remercie les enseignants de l'Université Toulouse 3 pour leur accueil et leur aide au cours de mes contrats d'enseignement, et notamment Laetitia Buisson, Christian Jost et Sergine Ponsard.

Un énorme merci à tous mes amis moulisiens, anciens ou récents, qui m'ont aidée et soutenue dans les meilleurs moments comme dans les pires. Merci à Anne-Sophie (A.) même si elle sera à l'autre bout du monde pendant ma soutenance. Merci à Nico (C.) pour son écoute et pour un dixième de ses blagues. Merci à Julien pour les barbecues au gîte. Merci à Elise pour les soirées mémorables au fin fond des montagnes. Merci à Alexis pour les crasses au bureau, les soirées jeux de société et le « uifi ». Merci à Léa pour les discussions féministo-politiques en salle thésards. Merci à Alice, reine burgonde devant l'éternel. Merci à Jérémie

pour avoir indéfectiblement meublé le silence de nos pauses clope. Merci à Robin de m'avoir changé les idées grâce aux rush b, aux bonnes bonnes et à la découpe de tomates en équipe. Merci à Ingrid de toujours dire ce qu'elle pense, et de toujours penser juste. Et n'oubliez pas qu'on a des soirées compotes à prévoir, tous les trois. Merci à Elvire pour ses conseils, sa gentillesse et ses câlins. Merci à Louis pour toutes ces heures passées à refaire le monde sur mon balcon (non loquetur). Et pêle-mêle, merci à Félix, Allan, Lauranne, Fia, Etienne, Kéoni, Aisha, Anne-Sophie (L.) ainsi que Ödil, Foxie et Bambou.

Merci aux amis (ex-)toulousains Kévin, Morgane, Valentin (si tu lis ces lignes, c'est qu'on s'en est sorti tous les deux) et plus particulièrement merci à Vincent et Marlène (et Izia) pour leur hospitalité et leur gentillesse. Merci à Myriam pour avoir été présente au pire moment. Merci aux GF pour la bouffée d'air frais qui permet de tenir les 362 jours restants. Un merci spécial à Nico (F.), pour les petites discussions rapides mais qui changent les idées et pour avoir rejoint le clan des silvers2 (ou 3 !). Et évidemment, je remercie Jade du fond du cœur pour avoir supporté mes angoisses, et pour être toujours là quand j'ai besoin d'elle.

Enfin, merci à ma famille pour son soutien indéfectible. Merci à Serge pour les bons petits plats et la rigolade. Merci à Florence pour les encouragements (et pour les lunettes !). Merci à Chloé pour m'avoir guidée lors de mes premiers pas dans la vie de thésarde. Merci à Louis pour m'avoir appris à ne pas me perdre en forêt. Merci à mes grands-mères pour m'avoir aidée et soutenue, chacune à leur façon. Merci à mon père pour les coups de fil nocturnes qui changent les idées et pour les billets de train salvateurs. Merci à Fabien et à Marion pour leur soutien, pour leurs visites et pour le ravitaillement, et merci pour le meilleur réveillon de tous les temps entre frangines. Enfin, merci à ma mère, pour m'avoir appris à admirer la nature, et pour tout le reste.

Table des matières

Liste des tables

Liste des figures

"He floated on his back when the valise filled and sank;
the river was mild and leisurely, going away from the people
who ate shadows for breakfast and steam for lunch and vapors for supper.
The river was very real; it held him confortably and gave him the time at last,
the leisure, to consider this month, this year, and a lifetime of years.
He listened to his heart slow. His thoughts stopped rushing with his blood."

Ray Bradbury, *Fahrenheit 451*

Introduction

Biodiversity: a multifaceted concept experiencing critical loss

Biodiversity is a concept referring to the variability of life on the planet (Ricklefs and Miller 2000). According to the Convention on Biological Diversity signed by 165 countries during the Rio Earth Summit in June 1992, biodiversity "includes diversity within species, between species, and of ecosystems".

Within-species diversity (hereafter named intraspecific diversity) is the most fundamental facet of biodiversity. It comprises genetic diversity (i.e. the diversity in the genetic information among individuals of a species) and phenotypic diversity (i.e. the diversity of behaviours, morphs and physiological traits among individuals of a species). Intraspecific diversity is the raw material on which acts selection, enabling species to adapt to environmental changes and ultimately leading to speciation (Ridley 2003). It improves species and communities resilience to disturbances (Jung et al. 2013; Moran et al. 2015). It plays a key role for evolutionary and ecological dynamics (Bolnick et al. 2003; Odling-Smee et al. 2003) by affecting the way species modulate their biotic and abiotic environment (Hughes et al. 2008; Bolnick et al. 2011).

Between-species diversity (hereafter named interspecific diversity) refers to the variety of species inhabiting an ecosystem and is the most obvious facet of biodiversity. Since the early 1990s, interspecific diversity has been hypothesized to positively influence ecosystem processes such as primary productivity, as well as their stability and maintenance in the face of disturbances (Loreau 2000*a*). Species richness (i.e. the number of species in an ecosystem) is extensively employed as a surrogate for interspecific diversity (Loreau 2010; Eduardo 2016) although functional complementarity among species is more and more acknowledged to have a greater impact on ecosystem functioning than species diversity per se (Eduardo 2016).

Finally, diversity of ecosystems refers to the number of different ecosystems on the planet. This notion accounts for the variety of unique ecological and evolutionary processes occurring in ecosystems such as interactions between species (predation, parasitism, coadaptation) and between species and their abiotic environment (primary production, nutrient cycling) (Begon et al. 2006).

Biodiversity can also be decomposed into three components: within-site diversity (α-diversity), between-site diversity (β-diversity) and the total regional diversity (γ-diversity) (Whittaker 1972). As an example, at the intraspecific level, the α component of biodiversity is the diversity found within a population (e.g. the genetic diversity displayed by the individuals of a same population) whereas the β component of biodiversity corresponds to the differentiation observed among populations (e.g. the genetic differentiation between two populations). Similarly, at the interspecific level, the α component of biodiversity is the diversity of species found within a community whereas the β component of biodiversity is the difference of species composition among communities. The γ component of biodiversity describes the overall intraspecific or interspecific diversity of all the populations or communities of a region and has been defined as being the product ($\gamma = \alpha \times \beta$; Whittaker 1972; Baselga 2010) or the addition ($\gamma = \alpha + \beta$; Lande 1996; Loreau 2000*b*) of the two other components.

All the facets of biodiversity are currently experiencing a dramatic decline caused by human actions (Butchart et al. 2010). This loss has been proven to directly affect ecological and evolutionary processes, thus reducing primary productivity and increasing ecosystems vulnerability to disturbances (Hooper et al. 2012). Additionally, the decrease in biodiversity has been shown to directly alter the benefits and services that humanity obtains from it, thus impacting human well-being (Díaz et al. 2006; Cardinale et al. 2012). In this context, the main challenge faced by scientists is to propose to decision-makers efficient and sustainable plans for limiting biodiversity loss. However, this requires an extensive understanding of the many facets and components of biodiversity, and notably of (i) how biodiversity is spatially distributed, (ii) what evolutionary and ecological processes shape the distribution of the facets and components of biodiversity and (iii) how the different facets of biodiversity interact with each other.

The patterns and drivers of biodiversity

Biodiversity is not evenly distributed on the planet (Gaston 2000). Some areas such as moist tropical forests harbour high biodiversity while others such as deserts are almost devoid of life. Studying this spatial distribution has allowed scientists to uncover repeated patterns

whereby biodiversity follows geographical or ecological gradients (Lawton 1996). For example, the most well defined and well known biodiversity pattern is the latitudinal gradient in interspecific diversity, i.e. the decrease of interspecific diversity from the equator to the poles (Pianka 1966). This pattern has been shown to hold true across many taxa, habitat types and geographic regions (Hillebrand 2004), and to affect both the α and the β components of interspecific diversity (Willig et al. 2003). Another example is the species-area relationship, whereby species richness in an ecosystem tends to increase with the area of the ecosystem (Connor and McCoy 1979).

The underlying drivers of these patterns are complex and can stem from diverse evolutionary and ecological processes involving biotic and abiotic features. As an example, the latitudinal gradient in interspecific diversity has been hypothesized to result from higher rates of speciation in the tropics (Mittelbach et al. 2007), as well as from a variation in predation pressure promoting species coexistence (Freestone et al. 2011). For this reason, understanding the origins of biodiversity patterns is still an ongoing challenge. Nonetheless, it is also a fundamental goal of evolutionary biology and ecology (Levin 1992; Ricklefs 2004; Chave 2013) for several reasons. First, it allows to better understand the functioning of ecosystems at different spatial scales (Gotelli et al. 2009). Second, it gives the possibility to make predictions about the repartition of biodiversity across landscapes and about its response to environmental changes (Guisan and Thuiller 2005). Third, as stated above, it helps establishing efficient conservation measures through the identification of the most diverse areas and of the key features sustaining biodiversity.

Studying biodiversity patterns within integrative frameworks

Despite this critical importance, most of the biodiversity patterns described to this day relate to interspecific diversity, notably because it is one of the most convenient proxy measures of biodiversity (Maclaurin and Sterelny 2008). However, the other facets of biodiversity and notably intraspecific diversity ought not to be neglected as they are of critical importance for ecosystem functioning, stability and resilence (Blanchet et al. 2017; Mimura et al. 2017).

Studying each facet of biodiversity independently does not appear sensible. Indeed, it has long been acknowledged that inter- and intraspecific diversity were driven by parallel processes (i.e. speciation/mutation, ecological/genetic drift, dispersal/gene flow, environmental filtering/natural selection) possibly shaping them in comparable ways (Antonovics 1976). Consequently, it has been hypothesized that inter- and intraspecific diversity patterns might be similar and this result has been verified empirically (as an example, Adams and Hadly (2013) showed that, like interspecific diversity, intraspecific genetic diversity followed a latitudinal gradient in numerous vertebrate species). Additionally, similar patterns of inter- and intraspecific diversity could ensue from direct interactions between the two levels of biodiversity (Vellend and Geber 2005). For instance, a high interspecific diversity might generate diversifying selection in a population through disparate predation, thus increasing the intraspecific genetic diversity within this population.

In this context, studying in a single framework inter- and intraspecific biodiversity patterns, their drivers and the possible interactions between them appears crucial to test these fundamental hypotheses. To that aim, Vellend (2003) introduced the Species-Genetic Diversity Correlation concept (SGDC, see Chapter II) which quantifies the congruency in the patterns of interspecific diversity and intraspecific genetic diversity. The studies on SGDC led to a better understanding of the relationships between inter- and intraspecific genetic diversity, as well as the processes shaping these facets of biodiversity in similar or contrasting ways (Taberlet et al. 2012; Vellend et al. 2014).

Testing for parallel patterns in other facets of biodiversity appears of great interest. For instance, studying the correlation between genetic intraspecific diversity and phenotypic intraspecific diversity within a framework similar to SGDC seems judicious as these two facets of diversity are intrinsically related and can be under the influence of similar adaptive and neutral processes (Lowe et al. 2017). For instance, in the case of neutral genetic markers and adaptive traits that are not strongly affected by selection, genetic and phenotypic intraspecific diversity are expected to display similar patterns if they are driven by neutral processes such as drift. Direct relationships between genetic and phenotypic diversity can also be expected, notably when genetic diversity directly codes for the considered phenotypic traits or appropriately describes the whole genomic diversity (Hoffman et al. 2014). Therefore, these facets of biodiversity are expected to covary under specific conditions, and

uncovering similarities or dissimilarities in their distributions should ensure a better understanding of the underlying processes. To that aim, I developed a new concept named Genetic-Phenotypic Intraspecific Diversity Correlation (GPIDC, see Chapter III).

Such integrative frameworks appear essential to get an acute appraisal of the distribution of biodiversity on the planet. Additionally, knowing to what extend and under which conditions several facets of biodiversity might covary would favour the establishment of conservation measures targeting areas sustaining high diversity at the inter- and at the intraspecific level. However, describing biodiversity patterns and understanding what drives them is a major statistical challenge as it implies disentangling intricate causal relationships among numerous factors. This issue calls for the use of appropriate statistical approaches.

Causal modeling as a method to study biodiversity patterns

Most of the approaches used for linking biodiversity descriptors (e.g. species richness, allelic richness, etc.) to environmental variables rely on empirical correlations. However, as "correlation does not imply (direct) causation" (Shipley 2000*a*), correlation provides no information on the underlying processes leading to the observed result. As an example, it is obvious that, even if a strong correlation is observed between latitude and biodiversity, latitude per se is not the direct cause of the latitudinal gradient in biodiversity. Consequently, the relations linking biodiversity descriptors to environmental variables have to be carefully interpreted, and may remain unexplained due to a statistical inability to tease apart causal relationships between variables (Shipley 2000*a*). Additionally, studying the patterns of several biodiversity facets in a single integrative framework such as SGDC calls for specific statistical approaches. Indeed, highlighting the environmental variables affecting two biodiversity descriptors while simultaneously testing for possible direct relationships between them requires statistical methods in which several dependent variables (here the biodiversity descriptors) can be taken into account simultaneously.

A solution to clarify entangled causal relationships within complex datasets may build on methods of causal modeling. Causal modeling assesses the validity of a model describing the expected causal links among a set of variables and estimates the strength of these links

(Grace 2006). Therefore, it allows to test for direct and indirect relations among diverse environmental variables and one or several biodiversity descriptors within a single framework. Two of the most used causal modeling approaches are maximum-likelihood based path analysis, initially introduced by Sewall Wright (1934) and the d-sep test developed by Bill Shipley (2000*b*). These methods are perfectly well suited to study biodiversity patterns at the α-level. However, at the β-level, the variables are in the form of pairwise matrices (e.g. matrix of genetic differentiation among sites, see above). The statistical analysis of pairwise matrices poses a series of analytical issues, notably because of the non-independence of pairwise data (Legendre and Legendre 2012; Graves et al. 2013) that prevents the application of causal modeling. Consequently, the first chapter of the present work has been dedicated to the development of novel statistical approaches allowing the application of maximum-likelihood based path analysis and of the d-sep test to data in the form of pairwise matrices (see Chapter I). In the two subsequent chapters (Chapter II and III), I used causal modeling to uncover common patterns in three facets of biodiversity (namely interspecific diversity and intraspecific genetic diversity in Chapter II and intraspecific genetic diversity and intraspecific phenotypic diversity in Chapter III) in the particular context of riverine networks.

Biodiversity patterns in riverine networks

Freshwater habitats cover approximately 0.8% of the Earth's surface but harbour 125,000 species of freshwater animals, representing 9.5% of all known animal species on the planet. This high concentration of biodiversity led some authors to define the entirety of freshwater habitats as a hotspot for biodiversity (Dudgeon et al. 2006; Strayer and Dudgeon 2010). This high interspecific diversity might be partially explained by the insular nature of freshwater habitats by which individuals are often restricted to a single biota (e.g. a lake or a river drainage), thus favouring speciation and endemism (Reyjol et al. 2007; Strayer and Dudgeon 2010). On the downside, the isolation between freshwater biotas impedes the ability of freshwater species to re-establish extinct populations and thus makes them very vulnerable to disturbances. Additionally, climate change is expected to strongly affect freshwater biodiversity (Heino et al. 2009) and human pressures on freshwater habitats are increasing, notably due to the rapid growth of the human population (Dudgeon et al. 2006).

Approximately 65% of all freshwater habitats have been declared to be under moderate to high threat due to human impacts such as water pollution, introduction of exotic species (e.g. Nile perch, water hyacinth, spinycheek crayfish) or flow modifications (Dudgeon et al. 2006; Vörösmarty et al. 2010). Remarkably, rivers are strongly impacted, with only an extremely small portion of the planet's rivers remaining unaffected by humans (Vörösmarty et al. 2010) while they harbour a high portion of the total freshwater biodiversity (Williams et al. 2004). Improving our understanding of the biodiversity patterns and drivers in riverine networks thus appears of critical importance to optimize their conservation.

Riverine networks are characterized by unique features. First, riverine networks harbour a distinctive dendritic structure (earning them the name of dendritic river networks, Fagan 2002; Grant et al. 2007) that consists of a mainstem to which are connected branches whose number typically increases from downstream to upstream. Second, riverine networks exhibit a strong upstream-downstream gradient of habitat capacities, with small physical carrying capacity in upstream branches (due to small river width) and larger carrying capacity in downstream branches and mainstems (larger river width). Third, riverine networks are characterized by an upstream-downstream gradient of environmental conditions. Upstream habitats are characterized by cold, highly oxygenated water, high water velocity, a coarse grain substrate and important vegetation cover while downstream habitats are characterized by warm and poorly oxygenated water, low water velocity, a fine grain substrate and scarce vegetation cover (Vannote et al. 1980).

Consequently, riverine networks are highly spatially-structured habitats, and biodiversity distribution is expected to highly contrast with what is observed within traditional two-dimensional landscapes (Altermatt 2013, see Chapter IV). The spatial connectivity of dendritic structures imposes huge dispersal constraints to freshwater species, especially in the case of species that can only disperse along the network (Grant et al. 2007, 2010). Notably, Carrara et al. (2012) experimentally showed a positive effect of dendritic structure on interspecific diversity. At the intraspecific level, Paz-Vinas and Blanchet (2015) theoretically demonstrated that dendritic configuration increased intraspecific genetic diversity when compared to two-dimensional lattice configuration. Additionally, migration in dendritic network is expected to be downstream-biased due to asymmetric dispersal costs caused by water flow and altitude (Morrissey and de Kerckhove 2009; Paz-Vinas et al. 2013) and the

combined effects of dendritic structure and asymmetric dispersal have been theoretically studied. For example, at the interspecific level, Muneepeerakul et al. (2008) found that interspecific β-diversity was increased by asymmetric dispersal in dendritic networks while, at the intraspecific level, Morrissey and de Kerckhove (2009) showed that dendritic structure and asymmetric migration could maintain higher levels of intraspecific genetic diversity. The spatial structuring in habitat capacities is also expected to influence evolutionary and ecological processes in riverine networks. At the interspecific level, Carrara et al. (2014) found a positive effect of spatially-structured habitat sizes on the evenness of community composition within a theoretical dendritic network when compared to randomly distributed or uniform habitat sizes. At the intraspecific level, Paz-Vinas et al. (2015) showed that the increase in habitat capacity downstream could generate an increase in genetic diversity. The impact of the environmental gradient on biodiversity has also been studied, mainly empirically. Notably, at the interspecific level, species assemblages are known to greatly vary along this gradient (Vannote et al. 1980; Schlosser 1990) and similarly, at the intraspecific level, individual phenotypes have been shown to follow this gradient (e.g. Hopper et al. 2015). These variations are expected to lead to high inter- and intraspecific β-diversity within riverine networks.

Disentangling the relative impacts of these features on biodiversity distribution appears challenging but is essential to the understanding and optimal protection of riverine networks. In complement to the empirical studies presented in Chapters II and III, I chose to use an eco-evolutionary model to theoretically solve this issue (Chapter IV).

Objectives

The main objective of the present work has been to study the patterns of inter- and intraspecific diversity within riverine networks. More specifically, I aimed at uncovering if interspecific diversity, intraspecific genetic diversity and intraspecific phenotypic diversity were similarly distributed, at both the α and the β levels. Additionally, I investigated which features of the riverine networks shaped interspecific diversity, intraspecific genetic diversity and intraspecific phenotypic diversity. This was achieved empirically through the use of integrative frameworks and theoretically through the use of an eco-evolutionary metapopulation dynamics model (Hanski et al. 2011).

This thesis has been divided into four chapters.

The first chapter has been dedicated to the development of statistical approaches allowing the application of causal modeling methods (namely path analysis and the d-sep test) to data in the form of pairwise matrices, in order to study biodiversity patterns and their underlying processes at the α and β levels within similar statistical frameworks.

In the second chapter, I studied the similarities and dissimilarities between interspecific diversity and intraspecific neutral genetic diversity patterns in four freshwater fish species (namely *Barbatula barbatula*, *Gobio occitaniae*, *Phoxinus phoxinus* and *Squalius cephalus*) within the Garonne-Dordogne riverine network. I investigated the causes of these patterns using an integrative framework (namely the SGDC framework, Vellend and Geber 2005) in order to uncover the processes underlying the spatial distribution of inter- and intraspecific diversity while taking into account the possible interactions between them.

In the third chapter, I tested for congruencies between the distributions of intraspecific neutral genetic diversity and intraspecific phenotypic diversity in two freshwater fish species (namely *G. occitaniae* and *P. phoxinus*) within the Garonne-Dordogne riverine network. I again used an integrative framework (named GPIDC) to uncover the drivers of these two facets of intraspecific diversity and the possible interactions between.

In the fourth and last chapter of this thesis, I used an eco-evolutionary metapopulation dynamics model to theoretically assess the impacts of the main features of riverine networks (i.e. the dendritic structure, the upstream-downstream gradient in habitat capacities, the upstream-downstream gradient in environmental conditions and the asymmetric dispersal rate caused by water flow) on local adaptation and on the distribution of intraspecific phenotypic diversity.

Chapter I - Inferring causalities in landscape genetics: An extension of Wright's causal modeling to distance matrices.

By Lisa Fourtune[1], Jérôme Prunier[1], Ivan Paz-Vinas[2,3,4], Géraldine Loot[1,5], Charlotte Veyssière[5] and Simon Blanchet[1,4]

[1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS); UMR 5321 (Station d'Écologie Théorique et Expérimentale, SETE), 09200 Moulis, France
[2] Aix-Marseille Université, CNRS, Institut de Recherche pour le Développement (IRD), Avignon Université; UMR 7263 (Institut Méditerranéen de la Biodiversité et d'Ecologie marine et continentale, IMBE), Centre Saint-Charles, Case 36, Marseille, France
[3] Université de Lyon, CNRS, École nationale des travaux publics de l'Etat (ENTPE); UMR 5023 (Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, LEHNA), 6 rue Raphaël Dubois, 69622 Villeurbanne, France
[4] CNRS, UPS, École Nationale de Formation Agronomique (ENFA); UMR 5174 (Laboratoire Évolution & Diversité Biologique, EDB), 31062 Toulouse cedex 4, France
[5] Université de Toulouse, UPS; UMR 5174 (EDB), 31062 Toulouse cedex 4, France

Accepted with minor revisions for publication in **The American Naturalist**

I.1 - Résumé

Identifier les caractéristiques du paysage qui affectent la connectivité fonctionnelle entre populations est un défi crucial tant pour l'écologie fondamentale qu'appliquée. La génétique du paysage combine des données génétiques et écopaysagères afin de résoudre cette problématique, et doit pour cela identifier les relations directes et indirectes existant au sein de jeux de données complexes. Le recours aux outils d'analyses causales (« causal modeling ») apparait dans ce contexte comme particulièrement adapté. Néanmoins, cette approche statistique n'a pas été initialement développée pour être appliquée à des données prenant la forme de matrices de distance, comme c'est souvent le cas en génétique du paysage. Dans cette étude, notre objectif est d'étendre le domaine d'application de deux méthodes d'analyses causales (le *path analysis* et le *d-sep test*) en développant des approches statistiques adaptées aux matrices de distances. Grâce à des simulations, nous avons démontré que ces nouvelles approches amélioraient grandement la robustesse des ces méthodes. A partir d'un jeu de données empiriques combinant des données génétiques d'une espèce de poisson d'eau douce (*Gobio occitaniae*) et des données écopaysagères, nous avons démontré l'intérêt des analyses causales pour mieux connaitre la connectivité fonctionnelle au sein de populations sauvages. Nous avons notamment démontré que des relations directes et indirectes impliquant l'altitude, la température et la concentration en oxygène de l'eau influençaient la diversité génétique inter- et intra-populationnelle chez *G. occitaniae*.

I.2 - Abstract

Identifying landscape features that affect functional connectivity among populations is a major challenge in fundamental and applied sciences. Landscape genetics combines landscape and genetic data to address this issue, with the main objective of disentangling direct and indirect relationships among an intricate set of variables. Causal modeling has strong potential to address the complex nature of landscape genetic datasets. However, this statistical approach was not initially developed to address the pairwise distance matrices commonly used in landscape genetics. Here, we aimed to extend the applicability of two causal modeling methods, i.e., maximum-likelihood path analysis and the directional-separation test, by developing statistical approaches aimed at handling distance matrices and improving functional connectivity inference. Using simulations, we showed that these approaches greatly improved the robustness of the absolute (using a frequentist approach) and relative (using an information-theoretic approach) fit of the tested models. We used an empirical dataset combining genetic information on a freshwater fish species (*Gobio occitaniae*) and detailed landscape descriptors to demonstrate the usefulness of causal modeling to identify functional connectivity in wild populations. Specifically, we demonstrated how direct and indirect relationships involving altitude, temperature and oxygen concentration influenced within- and between-population genetic diversity of *G. occitaniae*.

I.3 - Introduction

Landscape genetics is a discipline aimed at understanding spatial patterns of genetic diversity by exploring the relationships between landscape features and microevolutionary processes such as genetic drift, selection, mutation and gene flow (Manel et al. 2003; Manel and Holderegger 2013). This discipline builds on the latest advances in molecular biology and landscape data processing and is becoming increasingly important for fundamental and applied sciences (Storfer et al. 2010; Keller et al. 2015). Landscape genetics addresses issues ranging from the identification of barriers to dispersal, to the inference of the spread of non-native species (Storfer et al. 2010).

The main objectives of landscape genetics are to spatially describe effective dispersal (i.e., gene flow) and to identify landscape features (e.g., roads, dams, urban areas, and rivers) that affect functional connectivity (Manel et al. 2003; Storfer et al. 2010; Manel and Holderegger 2013). To achieve these objectives, landscape geneticists calculate genetic descriptors that are subsequently compared with landscape features and potential dispersal barriers (Balkenhol et al. 2009; Jaquiéry et al. 2011; Bradburd et al. 2013). Analytical tools developed for analyzing landscape genetic data often rely on empirical correlations that allow an assessment of the possible influence of various evolutionary processes. For example, a significant and positive correlation between genetic and geographic distances is generally considered indicative of isolation-by-distance (IBD: a spatial pattern whereby the homogenizing effect of gene flow decreases and the relative effect of genetic drift increases as the geographic distance between sites increases; Hutchison and Templeton 1999).

However, because “correlation does not imply causation”, processes can be incorrectly inferred from empirical correlations (Guillot et al. 2009). The likelihood of incorrectly inferring causalities from correlation is exacerbated in landscape genetics because it often implies intricate relationships among landscape variables. In the IBD example described above, the correlation between genetic and geographic distances might be direct, indirect and/or spurious. The correlation is “direct” (i.e., the migration rate between two sites decreases because they are far from one another) if no other variable co-varying with geographic distances causes the observed pattern of genetic distance. However, if a variable co-varies with geographic distances (e.g., the number of barriers between two sites) and causes the observed pattern, then the correlation between genetic and geographic distances is “indirect”. Alternatively a correlation is “spurious” when two variables are correlated because

they are both influenced by a third (unmeasured) variable (Cushman and Landguth 2010; Prunier et al. 2015). In the two latter cases, processes are incorrectly inferred from simple correlations. Consequently, the relationships linking landscape features to genetic descriptors have to be carefully interpreted, and they sometimes remain unexplained due to our inability to disentangle intricate relationships between variables (Shipley 2000*a*; Grace 2006). Clarifying causal relationships in landscape genetics is thus challenging, but important (Guillot et al. 2009).

A solution to improve inferences of causal relationships in landscape genetics may build on methods of causal modeling (e.g., Cushman et al. 2006). Causal modeling procedures, such as path analysis (Grace 2006), rely on the assessment of the validity of a causal graph describing the expected direct and indirect causal relationships among variables. Path analysis was initially developed by one of the founding fathers of population genetics, namely, Sewall Wright (1921). In path analysis, the influence along each path of the causal graph (i.e., the link between two variables) is estimated from correlation/covariance among the involved variables. Almost a century after its introduction by one of the most influential population geneticists, and despite its relevance for analyzing complex observational data, path analysis is still only occasionally used in landscape genetics and in population genetics in general.

Landscape geneticists generally focus on two main types of dependent variables that describe genetic diversity: (i) point summary statistics, which describe the genetic diversity at the sampling site level (e.g., allelic richness or heterozygosity) and (ii) pairwise summary statistics, which describe the genetic differentiation (or distance) between pairs of sampled populations or individuals (e.g., Fst, Jost's D). Several well-established methods allow a straightforward analysis of point summary statistics in a path analysis framework (Shipley 2000*a*; Grace 2006). For pairwise statistics, however, the process is more complex since the analysis of pairwise matrices poses a series of analytical issues, notably because of the non-independence of pairwise data (Legendre and Legendre 2012; Graves et al. 2013). Although pairwise data can be handled by reducing multidimensionality, using NMDS or dbRDA, for instance (e.g., Legendre and Fortin 2010), these types of analyses were not developed to tease apart direct and indirect relationships and are more suited to answer questions involving dissimilarity matrices rather than distance matrices (Legendre and Fortin 2010; Legendre et al. 2015). To address this specific data type, Cushman et al. (2006, 2013) proposed a causal modeling procedure based on partial Mantel tests to compare several competing causal models that link a matrix of genetic distances to matrices of explanatory variables. This

approach permits an assessment of the goodness-of-fit of each model by independently comparing the observed results of partial Mantel tests (partial correlation coefficients and associated p-values) to what is theoretically expected under each model specification. This approach has been proven to be powerful for inferring causalities from relatively simple models (Cushman and Landguth 2010). However, the design of the causal graph is constrained by the number of matrices of explanatory variables that can be handled in partial Mantel tests (only two), which limits the complexity of competing models and prevents the assessment of indirect relationships among variables. We believe that the use of alternative causal modeling procedures, such as maximum-likelihood-based path analysis (hereafter called "path analysis" for the sake of simplicity) and the directional-separation test (hereafter called "d-sep test", Shipley 2000*a*, 2000*b*), can represent an interesting improvement over the approach proposed by Cushman et al. (2006), as they may simultaneously account for all correlations implied in a model and permit the design (and comparison) of more complex models, explicitly addressing both direct and indirect effects.

We propose a simple and integrative framework to study direct and indirect links in the context of the analysis of landscape genetic data (and more generally, of ecological and evolutionary data involving pairwise matrices). As an introduction, we briefly present the philosophy, advantages and disadvantages of path analysis and the d-sep test. Then, we extend the applicability of these two methods to pairwise matrices (including distance and dissimilarity matrices) by developing two statistical approaches aimed at analyzing complex causal models (i.e., including several pairwise matrices linked both directly and indirectly) in landscape genetics. We then test the robustness of path analysis and the d-sep test applied to pairwise matrices using simulations. Finally, we use an empirical dataset involving patterns of genetic diversity in a freshwater fish species (*Gobio occitaniae*) and landscape descriptors at the river basin scale to demonstrate how these two statistical procedures can be used in landscape genetics to answer important biological questions. This study provides an opportunity to reconcile two important legacies of Sewall Wright's scientific life: population genetics and path analysis.

I.4 - A brief description of path analysis and the d-sep test

An introduction to causal graphs

Any causal modeling procedure is based on a causal graph illustrating the *a priori* hypotheses underlying the potential causal relationships within a set of variables. These relationships are depicted by vertices (i.e., nodes) representing variables that are linked by edges. A causal graph can contain *manifest variables* that are directly observed and measured (Shipley 2000*a*); *error variables*, which represent all of the factors that are not considered in the current graph; and *latent variables* that are hypothesized to exist but have not been measured directly (Grace 2006). Causal graphs are an intuitive approach to translate a causal hypothesis into a statistical language. The next step is to statistically test the relevance of the causal model in relation to data. Here, we focused on path analysis and the d-sep test, two methods dedicated to testing causal models without latent variables (Shipley 2000*a*; Grace 2006). These two methods are described below. When the causal graphs contain latent variables, the dedicated method is called structural equation modeling (SEM; Grace 2006), which is a generalization of path analysis. This method will not be presented here.

Path analysis

Path analysis is based on maximum likelihood estimation (Fisher 1950) of model parameters through the computation of covariance matrices. Each causal model includes a set of parameters, some of which are known (e.g., variances and covariances of variables), whereas others are unknown (e.g., path coefficients that quantify the direct influence of a variable along a given path; Wright 1921). The first step is to infer values for these unknown parameters. This inference is made iteratively by computing a maximum likelihood fitting function (F_{ML}; Bollen 1989) that quantifies the difference between the observed covariance matrix and a covariance matrix computed using the inferred values. The best parameter values are those that minimize this function. The absolute fit of the model can be assessed by computing a chi-square statistic and an associated p-value to determine whether the minimal value of F_{ML} is small enough to conclude that the observed data fit the hypothesized causal model; a high p-value indicates a high probability that the observed data fit the hypothesized causal model. Additionally, the relative fit of competing models can be tested using an information-theoretic approach (e.g., using Akaike's Information Criterion, AIC; Bollen 1989).

Path analysis requires linear relationships between variables, preferentially multivariate normal data (Shipley 2000*a*, Grace 2006), and assumes that observations are independent, which is notoriously not the case when considering pairwise matrices (Legendre and Legendre 2012). Because of this latter limitation, path analysis is not frequently used in landscape genetics.

D-sep test

Shipley's d-sep test simultaneously tests for conditional independence relationships that should be true if the causal model is verified. If these conditional independence relationships do not exist in the empirical data, then the causal hypothesis is rejected. These relationships are identified using the directional-separation (d-separation) criterion (Pearl and Verma 1987; Pearl 1988; Shipley 2003). However, there are usually far too many d-separation relationships to test all of them. The principle of the d-sep test is hence to identify a *basis set* of mutually independent d-separation relationships that together imply all others (Pearl 1988; Shipley 2000*b*). Once this basis set is identified, each of these k independence claims has to be tested against the empirical data. This can be achieved through the use of Pearson's partial correlation coefficients or linear regressions if the variables are normally distributed and are linked by linear relationships, as well as using more complex statistical methods in other cases. The k p-values obtained are equivalent to the probability levels of the data, given each of the k d-separation relationships. If all these tests are mutually independent, the k p-values can be combined using the following equation (Fisher 1938):

$$C = -2 \sum_{i=1}^{k} \ln(p_i) \quad (1)$$

If all the independence relationships hold in the data, this statistic follows a chi-square distribution with $2k$ degrees of freedom. The resulting test is called Fisher's C test (Shipley 2000*b*). A large C value, and thus a small resulting p-value, implies a poor absolute fit of the data to the model. In path analysis, the relative fit of competing models can also be assessed through the use of AIC adapted to the d-sep test (Cardon et al. 2011; Shipley 2013).

As the d-sep test does not impose any inference of parameters, the conditions for its application are flexible: it can, for instance, be applied to data sets with small sample sizes. Importantly, two nodes in a causal graph that are d-separated will also be conditionally independent in any dataset generated by this graph, irrespective of the distribution of the variables (Shipley 2000*a*; Pearl 2009). This means that different modeling approaches (e.g., linear or non-linear models, Bayesian models, hierarchical models, Shipley 2009; Cardon et al.

2011) can be used for testing the conditional independence relationships provided these tests are appropriate for the type of variables involved in the d-sep claims (Shipley 2009). The d-sep test is therefore a flexible method that cannot be used directly to infer path coefficients, although estimates can be computed from a combination of independent models.

A note about the use of p-values and AIC in the context of causal modeling

P-values and AIC provide different -yet complementary- information in the context of causal modeling. While AIC values allow the identification of the best fitting model among a set of candidate models (the relative fit), p-values provide information about the absolute fit of the empirical covariance matrix for a given model. This means that the best fitting model -with the lowest AIC- may be a poorly fitting model if diagnosed through the inspection of p-values. We therefore encourage users to use both the AIC and p-values to infer the causal structure of their data. On a philosophical side note, and following Goodman (1999), we here chose not to set any significance threshold (e.g., $\alpha = 0.05$): p-values are hereafter interpreted as the probability of obtaining a result equal to, or more extreme, than what was actually observed, under the null hypothesis.

I.5 - Extending path analysis and the d-sep test to pairwise matrices

We hereafter present four statistical approaches aimed at applying path analysis and the d-sep test to the analysis of causal models involving pairwise matrices. Fully usable R functions (R Development Core Team 2017) are presented in Appendix I-S1.

Path analysis applied to pairwise matrices

To take into account the non-independence of pairwise data in path analysis, we used the maximum likelihood population effects (MLPE) approach developed by Clarke et al. (2002) (see also Van Strien et al. 2012). In MLPE models, identities of the two sites involved in a pairwise comparison are treated as two random factors to take into account the spatial dependency of pairwise data: each site is associated with a random deviation from the intercept, and any pairwise values sharing a common source site thus share a common random deviation. To do so, we used the 'lavaan.survey' R package (Oberski 2014) that was initially developed to use SEM and path analysis with hierarchically structured data. In this approach

(hereafter named "clustering-based path analysis"), the identities of sites involved in a pairwise comparison were treated as clusters (i.e., each pairwise value was associated with two clusters), with the second cluster being nested within the first. As a result, all pairwise values originating from the same first cluster will share a common random deviation from the intercept, although each pairwise value will also be attributed a unique random effect associated with the second cluster being nested within the first. To prevent some sites from being more influential than others in the computation of the first random deviation, the identities of the two sites involved in a given pairwise comparison are randomly permuted. This approach allows the assessment of both the relative and absolute fits of a model, as well as the p-values associated with the path coefficients while partially taking into account the non-independence of pairwise data.

As an alternative approach to assess the p-values associated with path coefficients, we used a permutation procedure aimed at randomly permuting rows and columns of input matrices (Legendre 2000). This procedure provides the theoretical distribution of a given statistic (e.g., a Mantel correlation) under the null hypothesis of the absence of relationships between variables. An unbiased p-value can then be computed by comparing the observed value of the statistic to its null distribution. This approach (hereafter named "permutation-based path analysis") was not used to quantify the probability that the data fit the model, as the null distribution corresponds to a scenario in which none of the paths are true; rather, we aimed to test whether only the defined paths are true. This approach involves the following four steps. First, all matrices are independently permuted many times. Second, the values of the unknown parameters (i.e., path coefficients linking permuted matrices according to the considered causal model) are inferred by minimizing the difference between the covariance matrix computed from permuted data and the observed (optimized) covariance matrix to create a set of null causal models. The third and fourth steps consist of creating null distributions for the parameters of interest (here, values of the path coefficients) and computing unbiased p-values, respectively. The one-tailed p-value of each path coefficient is then computed as the proportion of permuted path coefficient values lower than or equal to (respectively greater than or equal to) the observed path coefficient value (Legendre 2000).

Finally, we built a parametric bootstrap procedure to quantify the range of values (i.e., confidence intervals) that act as good estimates of each unknown parameter value while taking into account the presence of pairwise matrices in the path analysis. This procedure is based on sampling with simultaneous replacement of rows and columns, performed numerous

times. Parameters values are estimated each time through path analysis, and 95% confidence intervals are provided as the 2.5% and 97.5% percentiles of these bootstrapped parameters. This method can only be applied to standardized data.

The d-sep test applied to pairwise matrices

Building on the flexibility of the d-sep test, we developed an approach that allowed the proper application of this method to distance matrices. This approach takes the form of a new R function called "*dsep.test*", permitting the use of the d-sep test for both point summary and pairwise data (see Appendix I-S1 for a user-friendly R script).

In the first step, the basis set of independent d-separation relationships implied by a given causal model is determined using the *basiSet* function from the R package 'ggm' (Marchetti 2006). In the next step, each conditional relationship of independence is tested using multiple regressions applied to distance matrices (MRM; Smouse et al. 1986), a permutation-based method classically used to infer parameters and p-values from regressions that can involve more than two pairwise matrices as explicative variables (contrary to the partial Mantel test). We used the *MRM* function from the R package '*ecodist*' (Goslee and Urban 2007). For each tested model, the p-values can then be obtained for each independence claim of the basis set and used as the p_i values in formula (1) to test the absolute fit of the model. This new approach is hereafter named "permutation-based d-sep test" for the sake of clarity.

In addition, we implemented an automatic calculation of the AIC score related to the tested model to test for the relative fit of competing models. AIC was only recently developed for the d-sep test (Cardon et al. 2011; Shipley 2013), and the R function has not been implemented in the 'ggm' package.

I.6 - Testing the reliability of path analysis and the d-sep test applied to pairwise matrices: a simulation test

General approach

To test the reliability of path analysis and the d-sep test to take into account the non-independence of pairwise data, we simulated 1000 datasets consisting of 50 sites each, and each site was associated with five observations that were independently drawn from five

normal distributions. These variables were separated into two independent variables (X_1 and X_2) and three response variables (X_3, X_4 and X_5). The response variables were calculated as linear combinations of one or two variables plus random noise (SD = 2). The causal model structure was held constant across simulations, but linear coefficients were randomly selected from a uniform distribution ranging from 0.8 to 1.6 for each simulation (see Appendix I-S2). From the five variables, we computed five Euclidean distance matrices (1225 pairwise values). With this procedure, distance matrices within simulated datasets were connected by four pre-defined causal links (Figure I-1*A*, Appendix I-S2). We then used these simulated datasets to test (i) the reliability of the p-values and AIC scores to detect adequate causal models among different model structures and (ii) the reliability of p-values and confidence intervals for specific path coefficients in the case of path analysis applied to pairwise matrices only.

First, we tested the ability of the new approaches to detect adequate causal models. We tested three types of models, each of which included four paths (Figure I-1*A*); the first type of model fitted the four causal relationships predefined in the simulated datasets (hereafter named "adequate model"), the second type of model fitted two of the four causal relationships (hereafter named "intermediate model"), and the third type of model fitted none of the four causal relationships (hereafter named "inadequate model"). For each of the 1000 simulated datasets, the three types of models were tested using the path analysis/d-sep test applied to point summary statistics, as well as the clustering-based path analysis and the permutation-based d-sep test applied to pairwise matrices. Clustering-based path analysis and the permutation-based d-sep test can be considered robust to assess the relative fit of each model if the AIC score of the "adequate model" calculated using these approaches is lower than the AIC scores calculated for the "intermediate" and "inadequate" models. We used the ΔAIC (the difference between the AIC of the considered model and the AIC of the best fitting model; Burnham and Anderson 2002) as a measure of the relative support of each model relative to the best fitting model (with $\Delta AIC < 2$ and < 4 as thresholds). Additionally, clustering-based path analysis and the permutation-based d-sep test can be considered more accurate than the approaches classically used for point-summary statistics to assess the absolute fit of the data to the model if the p-value computed for the "adequate model" is higher than the p-values computed using the classical approaches. For the "inadequate model", we expected both the classical approaches and the approaches developed for pairwise matrices to generate low p-values.

Second, we tested the ability of clustering-based path analysis and permutation-based path analysis to compute reliable p-values for the path coefficients of a given model, as well as the ability of the parametric bootstrap procedure developed for pairwise matrices to provide more reliable confidence intervals than bootstrap procedures not taking into account the non-independence of pairwise data. To do so, we built a model combining six paths (Figure I-2*A*): three of the paths fitted the causal relationships predefined in the simulated datasets (i.e., path coefficients of the "adequate model", hereafter named "adequate coefficients") whereas the other three did not (i.e., path coefficients from the "inadequate model", hereafter named "inadequate coefficients"). This model was fitted using classical path analysis, clustering-based path analysis and permutation-based path analysis, and confidence intervals were assessed using the parametric bootstrap procedure developed for pairwise matrices. The p-values of each path coefficient were obtained across the 1000 simulated datasets. In methods designed to account for the non-independence of pairwise data, the p-values of the "inadequate coefficients" should generally be high, whereas the p-values of the "adequate coefficients" should generally be low. On the contrary, the classical approach should consistently provide low p-values irrespective of the coefficient. Similarly, the 95% confidence intervals of "adequate coefficients" calculated using the parametric bootstrap procedure developed for pairwise matrices should not include zero, whereas the confidence intervals of "inadequate coefficients" should include zero.

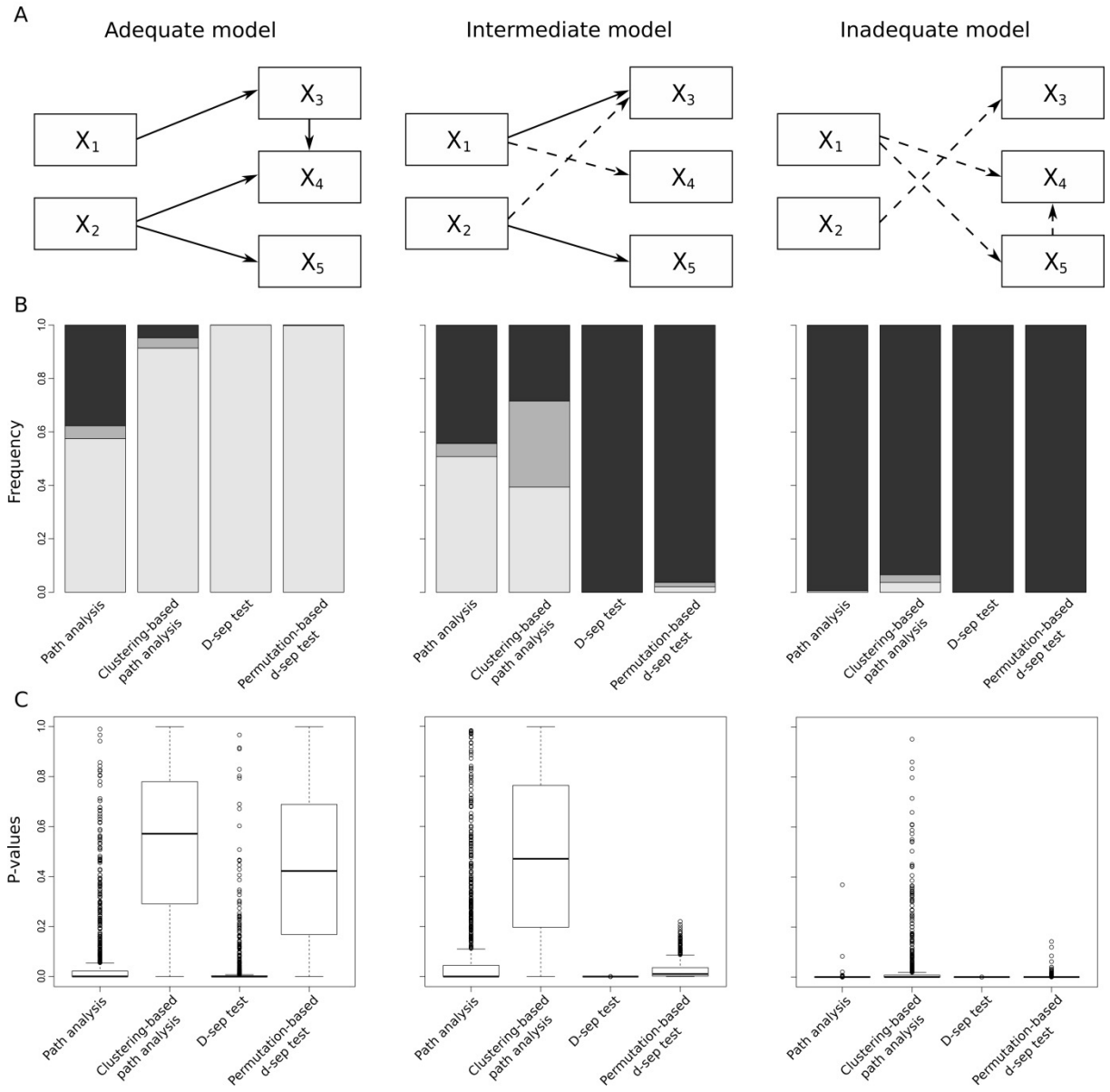


Figure I-1: A, Graphical representation of the "adequate", "intermediate" and "inadequate" models. B, Barplots summarizing the frequencies of ΔAIC values (dark grey: ΔAIC > 4, medium grey: 2 < ΔAIC < 4, light grey: ΔAIC < 2). C, Boxplots summarizing the p-values of both "adequate", "intermediate" and "inadequate" models obtained over 1000 simulations with classical path analysis, clustering-based path analysis, the classical d-sep test and the permutation-based d-sep test. The solid line within each box marks the median; the length of the box is the interquartile range (from the first to the third quartile). The lower whisker extends to the first quartile minus 1.5 times the interquartile range; the upper whisker extends to the third quartile plus 1.5 times the interquartile range. Small circles represent the data points which are beyond the whiskers.

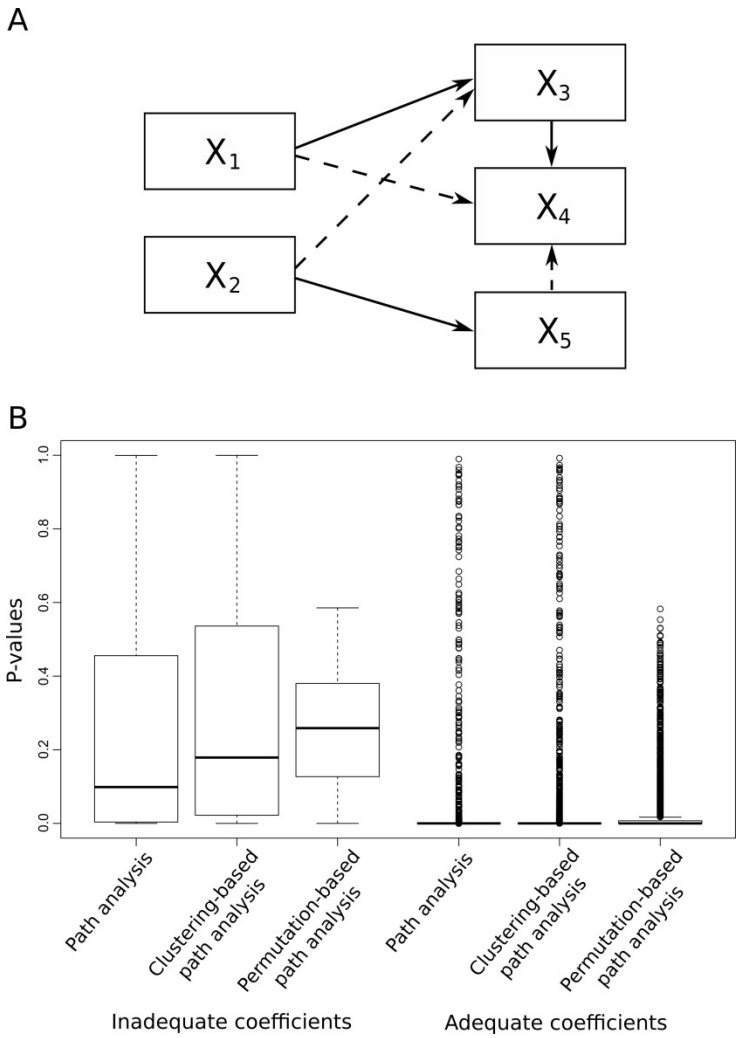


Figure I-2: A, Graphical representation of the model combining three "adequate" paths (black arrows) and three "inadequate" paths (dotted arrows). B, Boxplots summarizing the p-values of the "adequate" and "inadequate" coefficients obtained over 1000 simulations with classical path analysis, clustering-based path analysis and permutation-based path analysis. For the sake of clarity, the p-values of the three "adequate" and three "inadequate" coefficients are pulled together. See legend in Figure I-1C for details.

Results

AIC scores computed using clustering-based path analysis were more reliable than the AIC scores computed using classical path analysis to assess the relative fit of competing models (Figure I-1*B*). Notably, the "adequate model" was identified as one of the best fitting models in 91.4% of the simulations ($\Delta AIC < 2$, light grey bars in Figure I-1*B*) when using clustering-based path analysis, compared with only 57.5% of the simulations when using classical path analysis. Additionally, the "intermediate model" was identified as one of the best fitting models in only 39.4% of the simulations when using clustering-based path analysis, compared with 50.8% when using classical path analysis. However, ΔAIC of the

"intermediate model" ranged from 2 to 4 units in 32.2% of the simulations when using clustering-based path analysis (medium grey bars in Figure I-1*B*), indicating that this method penalized the absence of a part of the causal structure by only a small increase in the AIC score, and thus there was only a small decrease in relative fit. Interestingly, the relative fits of all three models were almost always correctly estimated using both the permutation-based d-sep test and the classical d-sep test, e.g., $\Delta AIC < 2$ was observed in 99.2% and 100% of the simulations, respectively, using the "adequate model" (dark grey bars in Figure I-1*B*).

P-values assessing the absolute fit of the "adequate model" were strikingly higher when clustering-based path analysis and the permutation-based d-sep test (median of the 1000 simulated p-values = 0.422 and 0.571, respectively) were used compared with the classical approaches (median < 0.001 for both the classical path analysis and d-sep test, Figure I-1*C*). This indicates that, when the null hypothesis is true, the p-value appears much more effective when taking into account the non-independence of pairwise data. Clustering-based path analysis and the permutation-based d-sep test hence appeared far more reliable than classical approaches to assess the absolute fit of a causal graph. The p-values of the "intermediate model" were high when clustering-based path analysis (median = 0.471) was used but were low when the permutation-based d-sep test (median = 0.010) was used, indicating a difference in the sensitivity of these two methods to detect models that do not perfectly reflect the causal structure underlying the data; the permutation-based d-sep test offers higher sensitivity than clustering-based path analysis. As expected, the p-values of the "inadequate model" computed using any of the approaches were very low (medians < 0.001 in all cases, Figure I-1*C*).

Clustering-based path analysis and permutation-based path analysis were also more reliable in estimating the p-values of path coefficients than classical path analysis (Figure I-2*B*). Indeed, accounting for the non-independence of pairwise data led to an increase in the p-values of inadequate coefficients (median of 0.179 and 0.259 with clustering-based path analysis and permutation-based path analysis, respectively; Figure I-2*B*), whereas the p-values of adequate coefficients remained low (median smaller than 0.001 for all methods). Additionally, the parametric bootstrap procedure developed for pairwise matrices greatly improved the reliability of the 95% confidence intervals of inadequate coefficients: only 4% of confidence intervals computed for inadequate coefficients did not include zero when taking into account the pairwise matrix structure of the data. However, the confidence intervals of adequate coefficients computed using this procedure included zero in 28.5% of the simulations.

I.7 - Empirical illustration of path analysis and the d-sep test

We used a dataset involving a freshwater fish species (the gudgeon *Gobio occitaniae*) to illustrate how path analysis and the d-sep test can be used when the causal models include either point summary statistics or pairwise matrices. Our aims were (i) to unravel direct and indirect relationships between landscape features (and associated processes) and genetic diversity in dendritic river networks and (ii) to disentangle the relative effects of natural vs. anthropogenic factors on spatial patterns of genetic diversity in gudgeon.

Data collection

Genetic data. A total of 92 sites scattered across the whole Garonne-Dordogne river catchment (southwestern France, see Figure I-3*A*) were sampled, and a maximum of 30 gudgeons per site were caught by electrofishing during spring 2010 and 2011. Nine out of these 92 sites were discarded from the analysis because fewer than ten samples were available. The final database included 1993 individuals sampled from 83 sites distributed across 34 rivers (see Appendix I-S3). For each individual, a small piece of pelvic fin was collected and preserved in 70% ethanol. DNA was extracted using a salt-extraction protocol (Aljanabi and Martinez 1997), and individuals were genotyped for eight microsatellite loci as described in Blanchet et al. (2010). Neither departure from Hardy-Weinberg equilibrium nor null alleles were detected for any of these loci (Fourtune et al. 2016). Eight samples were not successfully genotyped and were removed from the database.

Genetic diversity at the sampling site level was assessed using Fstat 2.9.3 (Goudet 2001) by computing the standardized allelic richness (Ar), i.e., the expected mean number of alleles (over all loci) in a random subsample of N individuals at each sampling location, where N is the smallest sample size across populations (N = 10). Genetic differentiation among sampling sites was measured using Jost's D (Jost 2008). This metric measures the allelic variation between pairs of populations; it has a null (or slightly negative) value when there is no differentiation between two populations and a value of one when two populations have no alleles in common. Jost's D among sites was calculated using the ‘mmod’ package (Winter 2012) in the R environment.

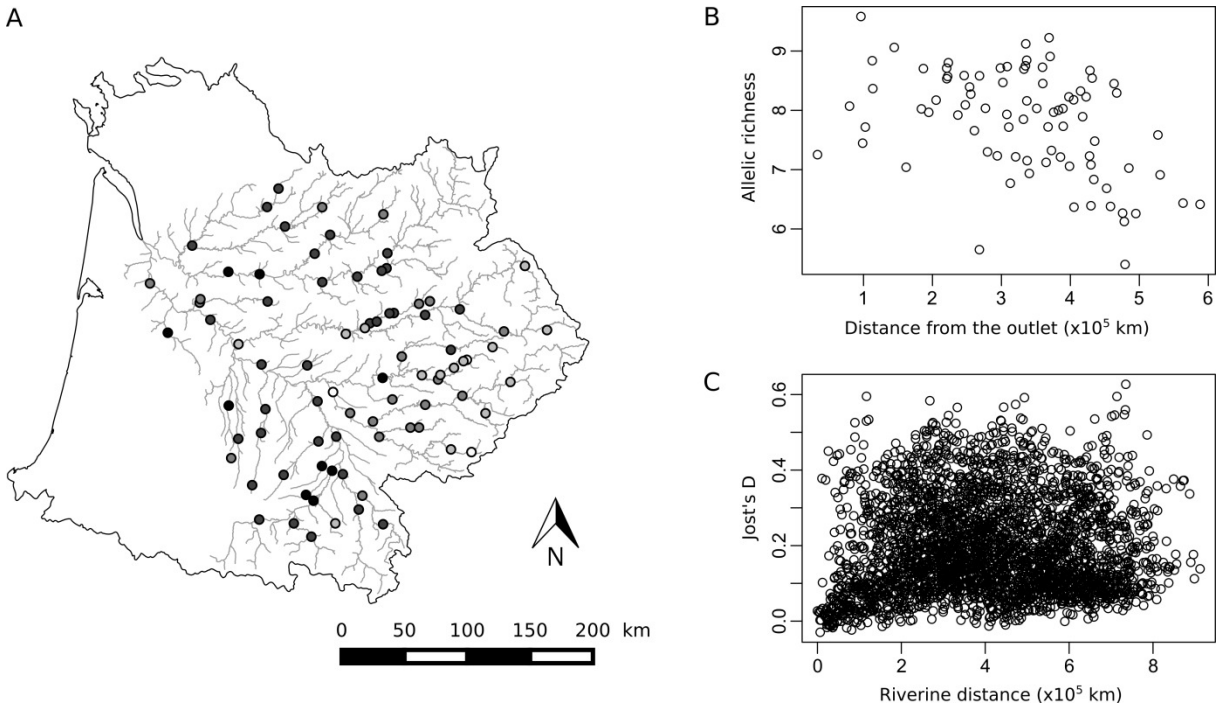


Figure I-3: Characteristics of genetic data. A, Location of the 83 sites, colored according to their allelic richness. White: lowest values, black: highest values. B, Allelic richness plotted against distance from the outlet. C, Jost's D plotted against pairwise riverine distance.

River topography. We selected three variables describing the topography and network arrangement at each sampling site, as network connectivity and topology are known to affect biodiversity patterns in river networks (Campbell Grant et al. 2007; Carrara et al. 2012; Paz-Vinas and Blanchet 2015). First, the betweenness centrality value (an index of river connectivity quantifying the positional importance of a node within a network; Freeman 1977) of the closest confluence upstream from each site was estimated using NetworkX (Hagberg et al. 2008), with higher values corresponding to nodes of higher importance for network connectivity. Second, local altitude and distance from the river mouth were obtained from the French Theoretical Hydrological Network (Réseau Hydrologique Théorique français, RHT; Pella et al. 2012). Third, the geographic distance along the river network (riverine distance) between each pair of sites was computed using QuantumGIS software (QGIS; Quantum GIS Development Team 2017).

Physico-chemical quality. We hypothesized that the physico-chemical characteristics of the sampling sites may affect the density of the fish populations (i.e., sampling sites with good water quality and optimal physical properties should sustain high fish densities) and hence, the effective population size (assuming a positive correlation between abundance and

effective population size and ultimately, genetic diversity). The data thus indirectly reflected the possible influence of genetic drift on the genetic summary statistics. Data were obtained from the database of the Water Information System of the Adour Garonne basin (Système d'Information sur l'Eau du Bassin Adour Garonne, SIEAG; http://adour-garonne.eaufrance.fr). Among other variables, this database compiles chemical characteristics of surface water (e.g., concentrations of various chemical compounds), measured several times a year at numerous sites in the Garonne-Dordogne river basin. Only sites with data available for March, May, July, September and November of 2011 were selected from the SIEAG database. Most of our sampling sites overlapped with a SIEAG site, in which case the mean of the five temporal measures was used as a proxy for the chemical quality of our sampling sites. When the overlap was not perfect, each sampling site was assigned to the nearest SIEAG site (on the same river), and the average values of variables at this nearest SIEAG site were used as surrogates for the chemical quality of the sampling site. Three sampling sites had no SIEAG site close enough to obtain reliable information (distance greater than 10 km) and were therefore discarded from the final database.

We specifically obtained water temperature and oxygen concentration data to test the assumption that a site with an optimal temperature and a high oxygen concentration can host larger fish densities. Additionally, we selected five chemical components directly affected by human activities and considered as good indicators of water quality (ammonium, nitrate, nitrite, orthophosphate and phosphorus concentrations). Principal component analysis (PCA) of these five chemical components (see Appendix I-S4) was performed using the R package 'ade4' (Dray and Dufour 2007). The coordinates of each site on the first axis, which accounted for 62.4% of the variance, were used to create a synthetic variable (hereafter called "chemicals"), representing the amount of chemical components at each site. Low values correspond to high ammonium, nitrate, nitrite, orthophosphate and phosphorus concentrations.

Habitat fragmentation. We selected three variables related to habitat fragmentation, as it has previously been shown to affect genetic diversity in gudgeon (Blanchet et al. 2010). Habitat fragmentation variables were obtained from the ROE database (Référentiel des Obstacles à l'Écoulement; ONEMA 2010) that identifies and georeferences barriers to water flow along French rivers. Two main types of obstacles, weirs (< 4 m high) and dams (5-30 m high in general), were considered here. We measured the "home-range" of each population, i.e., the riverine geographic distance a fish can access without being stopped (both upstream,

downstream and in tributaries) by a weir or dam (Prunier et al. 2017). This is thus a direct measure of the impact of weirs and dams on fish habitat availability. Furthermore, the total number of weirs and dams along the river stretch between each pair of sites was calculated. Because these obstacles may have differential influence on genetic diversity, they were considered separately.

Statistical analysis

Point summary statistics. Allelic richness (Ar) was analyzed using both classical path analysis and the d-sep test. For the two approaches, a full causal model was first designed using theoretical and *a priori* knowledge. In this model (Figure I-4*A*), allelic richness is expected to be directly linked to two human-related factors (chemicals and home range) and to five natural factors (betweenness centrality, oxygen concentration, temperature, distance from the mouth and altitude). We constrained altitude and distance from the mouth to be related one to the other, and we assumed direct but also indirect relationships between altitude and allelic richness through temperature and oxygen concentration (Figure I-4). All other relationships were direct (Figure I-4). Prior to analysis, all variables were centered and scaled to obtain standardized parameter estimates (Schielzeth 2010).

This full model was tested through path analysis using the *sem* function from the 'lavaan' R package (Rosseel 2012). We thus obtained the F_{ML} value, the corresponding p-value and the AIC of the model. We used the function dsep.test available in appendix part A to test the full model using a d-sep test approach in which conditional dependencies were tested through linear regressions. The full model was then simplified by removing paths one by one until the model with the best relative fit (i.e., the lowest AIC score) was identified. The absolute fit of this model was computed to ensure that the observed data were coherent with the model. This process permitted the identification of variables with a major influence on allelic richness. Finally, path analysis was used to collect the inferred path coefficients and the residual variance of allelic richness (corresponding to the amount of variance that was not explained by the model).

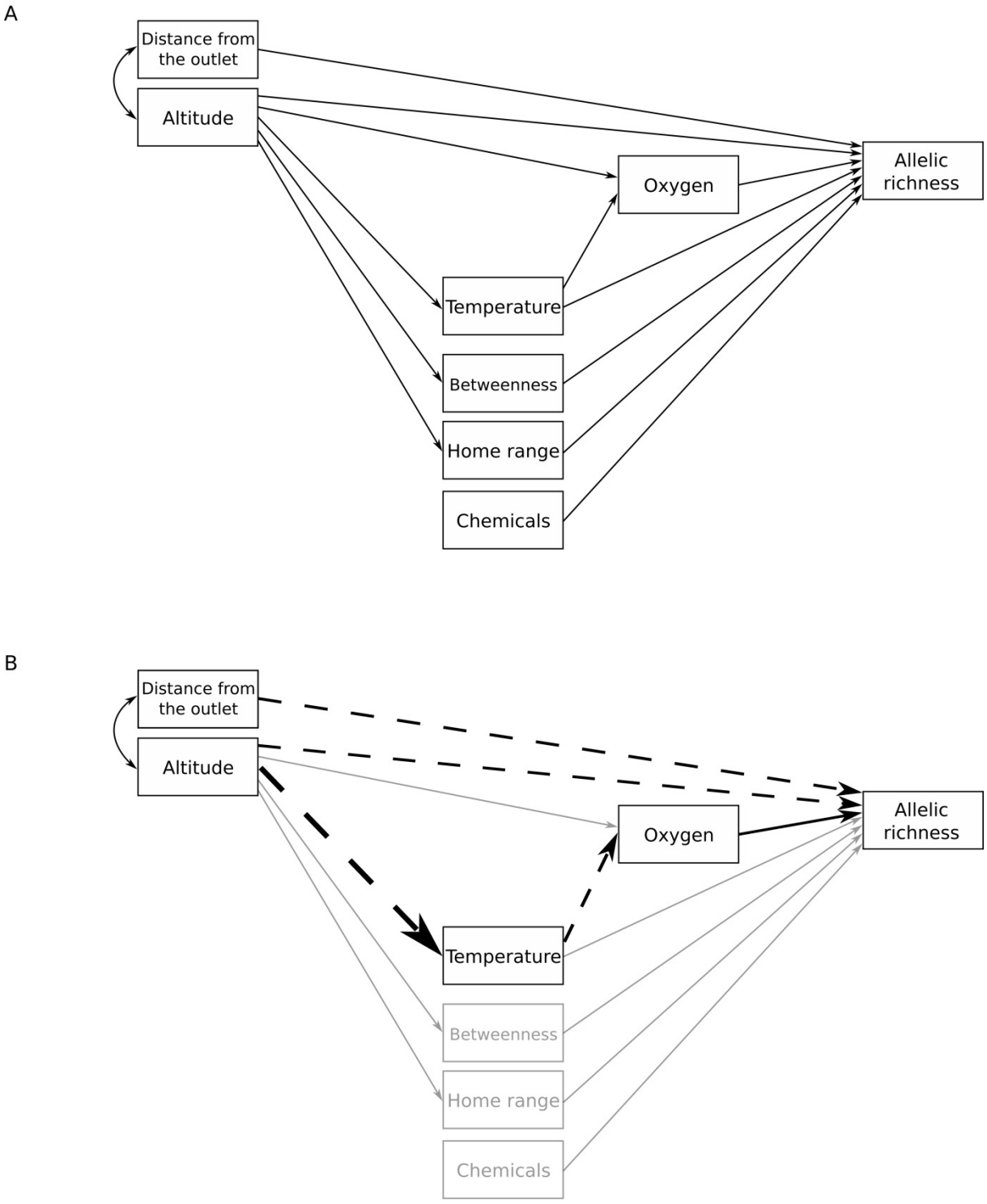


Figure I-4: Graphical representations of A, the complete model depicting causal relationships between allelic richness and anthropogenic and natural factors and B, the optimal model obtained after the simplification procedure. Single-headed arrow indicates a causal link. Double-headed arrow indicates co-variation. Solid and dashed lines stand for positive and negative values, respectively. Their width is proportional to the absolute value of the corresponding path coefficient. In grey, paths removed during the simplification procedure.

Pairwise summary statistics. In our dataset, four variables were present in the form of pairwise distance matrices: Jost's D (dependent variable), the counts of weirs and dams (separately) between each pair of sites, and the river distance between pairwise sites. Additionally, pairwise dissimilarity matrices for chemicals, altitude, temperature and oxygen concentration were computed as the absolute differences between sites. These dissimilarity matrices reflected the isolation-by-environment hypothesis (IBE; Rundle and Nosil 2005; Sexton et al. 2014). Using these eight variables, we designed a complete model (Figure I-5*A*) in which the number of weirs and dams between sites and pairwise differences in oxygen concentration and chemicals have a direct effect on genetic differentiation, whereas pairwise differences in altitude, temperature and riverine distance have both direct and indirect effects (Figure I-5*A*). As conducted previously, all variables were centered and scaled to facilitate interpretation. The full model was tested through both clustering-based path analysis and the permutation-based d-sep test. The model was simplified as explained previously until the model with the best relative fit was obtained. The coherence of the observed data to the best-fitted model was tested using clustering-based path analysis and the permutation-based d-sep test, whereas the p-values and confidence intervals of its path coefficients were computed using clustering-based path analysis, permutation-based path analysis and the parametric bootstrap procedure developed for pairwise matrices.

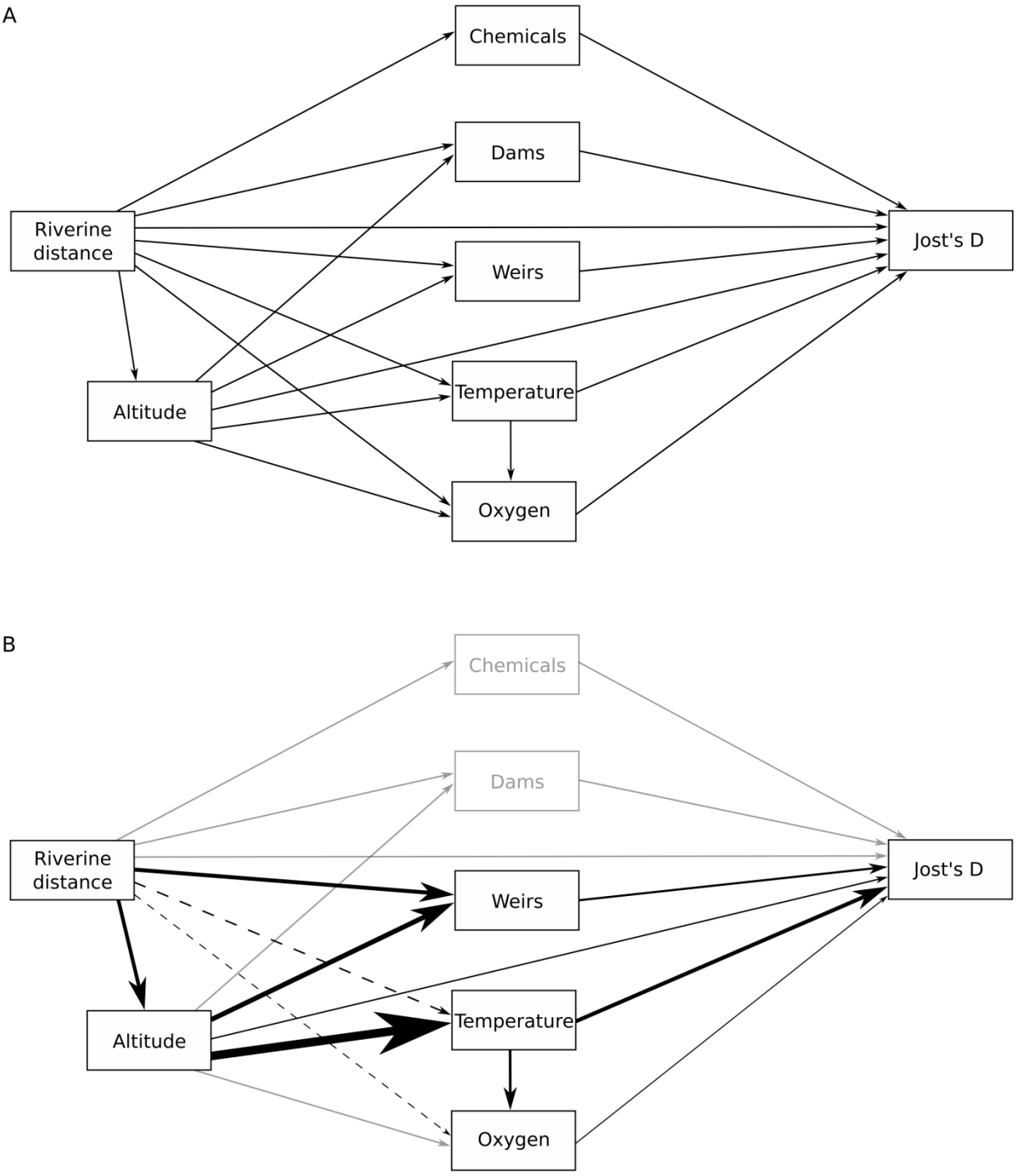


Figure I-5: Graphical representations of A, the complete model depicting causal relationships between genetic differentiation and anthropogenic and natural factors and B, the optimal model obtained after the simplification procedure. See legend in Figure I-4 for details.

Results

Description of genetic data. Ar ranged from 5.401 to 9.582, with a mean value of 7.793 (± 0.887). No apparent spatial pattern could be visually identified (Figure I-3*A*).

However, there was a significant decrease in allelic richness from downstream to upstream sections of the landscape (Pearson correlation between allelic richness and distance from the river mouth, r = -0.513, d.f. = 1, p < 0.001, Figure I-3*B*).

Jost's D ranged from -0.029 to 0.627, with a mean value of 0.196 (± 0.124). We failed to detect any IBD pattern since Jost's D did not increase with the riverine distance between sites (Mantel test, r = 0.055, p = 0.172, Figure I-3*C*).

Causal modeling applied to point summary statistics. Simplification of the full model led to the removal of seven paths before the model with the lowest AIC score was obtained (Table I-1 and Figure I-4*B*). Both approaches (path analysis and the d-sep test) led to the same best-fitted model. In this best-fitted model, allelic richness was directly correlated with the distance from the outlet, the altitude and the oxygen concentration. Allelic richness was higher in downstream sites at a low altitude and with a high oxygen concentration (Table I-2). Altitude was also indirectly correlated with allelic richness through a pathway that sequentially involved temperature and the oxygen concentration (Figure I-4*B*). The residual variance of Ar was 0.508 ± 0.080, which indicated that almost 50% of its total variance was explained by this best-fitted model.

Table I-1: Path analysis and d-sep test statistics used to disentangle the effects of environmental factors on allelic richness (point summary statistics) and genetic differentiation (pairwise summary statistics) using simplification procedures.

	Test statistics	d.f.	p-value	AIC
Path analysis on point summary statistics				
Complete model	41.455	15	<0.001	11.445
Optimal model	1.656	4	0.799	-6.344
Dsep test on point summary statistics				
Complete model	59.676	30	0.001	109.676
Optimal model	4.549	8	0.804	32.549
Clustering-based path analysis on pairwise summary statistics				
Complete model	82.273	10	<0.001	62.273
Optimal model	3.286	4	0.512	-4.714
Permutations-based dsep test on pairwise summary statistics				
Complete model	57.509	20	<0.001	121.509
Optimal model	4.149	8	0.843	46.149

Causal modeling applied to pairwise matrices. Irrespective of the approach, the simplification procedure led to the gradual removal of six paths before the model with the lowest AIC score was obtained (Table I-1 and Figure I-5*B*). In this best-fitted model, Jost's D was directly correlated with the differences in altitude, temperature, oxygen concentration and the number of weirs between sites. All these explanatory variables were also correlated with riverine distance through direct pathways and, in the case of weirs, temperature and oxygen, through indirect pathways involving other environmental variables (Table I-2, Figure I-5*B*).

Table I-2: Estimates of the path coefficients of the optimal models and their associated p-values and 95% confidence intervals (95% CI) obtained using path analysis. Clustering-based path analysis and permutation-based path analysis used on pairwise summary statistics provided similar p-values.

	Path coefficient	p-value	95% CI
Point summary statistics			
Distance from the outlet → Allelic richness	-0.407	0.005	[-0.720;-0.127]
Altitude → Allelic richness	-0.392	0.005	[-0.663;-0.095]
Oxygen → Allelic richness	0.377	<0.001	[0.227;0.554]
Temperature → Oxygen	-0.453	<0.001	[-0.648;-0.266]
Altitude → Temperature	-0.745	<0.001	[-0.899;-0.599]
Pairwise summary statistics			
Altitude → Jost's D	0.094	<0.001	[0.052;0.139]
Oxygen → Jost's D	0.071	<0.05	[0.039;0.105]
Temperature → Jost's D	0.265	<0.001	[0.226;0.304]
Weirs → Jost's D	0.144	<0.001	[0.110;0.182]
Riverine distance → Weirs	0.266	<0.001	[0.234;0.299]
Altitude → Weirs	0.329	<0.001	[0.289;0.369]
Riverine distance → Oxygen	-0.069	<0.001	[-0.101;-0.038]
Temperature → Oxygen	0.177	<0.001	[0.142;0.213]
Altitude → Temperature	0.577	<0.001	[0.544;0.607]
Riverine distance → Temperature	-0.104	<0.001	[-0.135;-0.072]
Riverine distance → Altitude	0.247	<0.001	[0.214;0.283]

I.8 - Discussion

Within a landscape, environmental variables can have compounding and contrasting impacts on spatial patterns of genetic diversity, and properly inferring these impacts is a key challenge in landscape genetics (Storfer et al. 2010). Here, we built on the framework of path analysis and the d-sep test (which we extended to the analysis of pairwise matrices) to provide

landscape geneticists with a reliable statistical tool to improve their ability to unravel direct and indirect relationships between landscape features and the spatial distribution of genetic diversity.

From the validation of path analysis and the d-sep test applied to pairwise matrices...

We improved the two commonly used causal modeling approaches (path analysis and the d-sep test; Shipley 2000*b*; Grace 2006) by extending their validity to the analysis of causal models comprising pairwise matrices and by using various procedures such as permutations, bootstrapping and specification of random effects. We provide operational R functions for these four improved approaches in Appendix I-S1, making them directly transferable to other biological systems. The simulations demonstrated that, as expected, our improved procedures were robust in identifying the best causal model compared to path analysis or d-sep tests that do not explicitly account for non-independence of pairwise data. Although clustering-based path analysis does not account for total non-independence between pairwise data because of the nested structure of random effects, it is noteworthy that the approach is a major improvement over classical path analysis. However, clustering-based path analysis imperfectly assessed the relative and absolute fits of models that did not perfectly reflect the causal structure underlying the data ("intermediate model"). We therefore suggest that clustering-based path analysis be used as a secondary approach to ensure/refine results obtained from the permutation-based d-sep test, if needed.

This simulation study was also used to test whether clustering-based path analysis and permutation-based path analysis were reliable in inferring the p-values of path coefficients connecting pairwise matrices and hence model parameters. We showed that both methods greatly outperformed traditional path analysis. Additionally, the parametric bootstrap procedure developed for pairwise matrices provided reliable confidence intervals for path coefficients, despite a tendency to underestimate these intervals in the case of adequate coefficients.

Although robust, the results of our simulation study should be considered with caution as we only explored a small set of variables. We therefore call for additional simulations (Landguth et al. 2015) to further assess the reliability of clustering-based path analysis, permutation-based path analysis and the permutation-based d-sep test, especially when compared to other traditional statistical procedures used in landscape genetics. Furthermore, we encourage further methodological developments for the implementation of the MLPE

procedure (Clarke et al. 2002) into the framework of clustering-based path analysis, as it is currently based on a hierarchical structure of random effects.

...To application in a real landscape.

When applied to empirical genetic data for *G. occitaniae* obtained from a whole river basin, both path analysis and the d-sep test identified the best-fitted causal models depicting both direct and indirect relationships between genetic summary statistics and landscape predictors. Our results strongly suggest that both allelic richness and pairwise measures of genetic differentiation were mainly related to natural landscape features (altitude, temperature and oxygen concentration), and, at such a large spatial scale, anthropogenic factors (related to habitat fragmentation and water pollution) were negligible drivers of genetic diversity in this species (with the exception of weirs).

Interestingly, some landscape features such as altitude were identified as direct drivers of both allelic richness and genetic differentiation. There was indeed a strong direct negative relationship between allelic richness and altitude, indicating that allelic richness was higher in sampling sites located at lower altitudes. Similarly, we found a direct positive relationship between the difference in altitude and pairwise measures of genetic differentiation between sites: the higher the difference in altitude, the higher the genetic differentiation. This type of direct relationship between altitude and genetic summary statistics for a freshwater fish is, to our knowledge, rarely presented in the literature (but see Faulks et al. 2011) and could reflect two non-exclusive processes: the past colonization history of *G. occitaniae* and the contemporary influence of asymmetric gene flow toward downstream sites. First, altitude may reflect historical contingencies whereby glacial refugia during the last glaciation event (~10 000 years ago) were mainly found in lowlands. Glacial refugia (in downstream sites) should indeed be more genetically diverse than recently colonized areas (in upstream sites; Paz-Vinas et al. 2015), whereas larger distances from the refugia should be associated with higher genetic differentiation (Costedoat and Gilles 2009). Second, altitude could be a good surrogate for the unidirectional water flow of rivers that favors gene flow from upstream (sites at high altitude) to downstream (sites at low altitude; Paz-Vinas et al. 2015). The direct negative relationship between distance from the outlet and allelic richness could similarly stem from these two mechanisms.

For both allelic richness and pairwise measures of genetic differentiation, we additionally found indirect relationships between altitude and genetic diversity. These indirect

relationships were mediated through water temperature and oxygen concentration. These two indirect relationships were expected to underline the effect of genetic drift. Oxygen is an important driver of fish species distribution in river networks (Crispo and Chapman 2008), and we hypothesized that higher oxygen concentrations may sustain higher fish densities (or at least sites with extremely low oxygen availability may have higher fish mortality). Assuming that density is positively related to the effective population size in fish (Belmar-Lucero et al. 2012), oxygen limitation may directly alter allelic richness and ultimately, genetic differentiation through genetic drift (Hutchison and Templeton 1999). In the same way, because water temperature and oxygen availability are negatively correlated, higher water temperatures may also be related to lower effective population sizes, leading to an increase in genetic drift and consequently, genetic differentiation. This unmeasured effect of genetic drift may also explain the direct effects of differences in altitude and water temperature on genetic differentiation. The direct links between differences in water temperature, differences in oxygen concentration and genetic differentiation may stem from the additional effect of IBE, a process that occurs when populations inhabiting different environments experience divergent patterns of selection (Rundle and Nosil 2005; Sexton et al. 2014). As a consequence, dispersing individuals may be maladapted to new environments, with reduced fitness and reproductive success and thus decreased gene flow between environmentally different areas (Crispo et al. 2006). Water temperature and oxygen concentration may thus act as important selective pressures in *G. occitaniae*, a hypothesis that deserves further investigation. Nonetheless, our study presents, to our knowledge, one of the first demonstrations of a direct relationship between oxygen availability, water temperature and genetic diversity in a freshwater fish species.

Regarding anthropogenic factors, weirs were found to have an impact on genetic differentiation, whereas the link between dams and genetic differentiation was discarded. This is surprising since weirs are not as high (1-4 m) as dams (5-30 m) and are generally expected to be more permeable to dispersal than dams (Blanchet et al. 2010). However, weirs are also generally older (they can be as old as 400 years, whereas dams are generally no more than 60 years old): considering the possible delay between anthropogenic impacts and the ensuing genetic response (Smith and Bernatchez 2008), it is possible that dams are too recent to have left a significant genetic imprint on the spatial patterns of genetic differentiation. A non-exclusive hypothesis may be that the relatively low influence of dams on genetic differentiation can be explained by their small numbers in the network (there are on average

fourfold fewer dams than weirs between sites). The influence of weirs as factors limiting dispersal and increasing genetic differentiation in freshwater fish has been shown in previous studies (Raeymaekers et al. 2009; Blanchet et al. 2010; Faulks et al. 2011), and our study therefore confirms these findings at a larger spatial scale while taking other co-variables into account.

It is noteworthy that we also found indirect relationships between geographic isolation and genetic differentiation (through the number of weirs and the differences in altitude, water temperature and oxygen concentration between sites), despite the absence of any direct relationship between riverine distance and Jost's D. IBD patterns are generally interpreted as imprints of gene flow and genetic drift (Hutchison and Templeton 1999), although the exact mechanisms underlying these patterns are rarely unraveled. Here, using path analyses, we were able to highlight potential causal pathways linking geographic distance to genetic differentiation: this relationship most probably arose from the spatial co-variation between geographic distances, number of weirs, and differences in altitude, water temperature and oxygen concentration. The use of causal modeling allowed the unraveling of multiple and complex relationships between geographic distance and genetic differentiation, which is essential for fundamental knowledge and applied perspectives. Of course, the relationship between geographic isolation and genetic differentiation may also result from alternative processes that we failed to model properly (Ewers and Didham 2006), and this should be investigated in future studies. For instance, the complexity of the river network is expected to play a major role in genetic differentiation of aquatic organisms (Paz-Vinas and Blanchet 2015), with isolated upstream populations acting as reservoirs for unique and rare alleles, hence triggering high genetic differentiation between upstream and other populations.

Causal modeling taking both direct and indirect effects into account provides a better appraisal of factors driving the spatial distribution of alleles in river networks. Most previous studies on spatial patterns of genetic diversity in aquatic organisms focused on the relationships between allelic richness and distance from the river mouth and generally found an increase in genetic diversity from the source to the mouth of river networks (reviewed in Paz-Vinas et al. 2015). We also uncovered this general pattern, although the use of path analysis procedures suggested that several other environmental variables were linked to allelic richness, even in a model in which the relationship between distance from the outlet and allelic richness was taken into account. This result illustrates the strength of causal modeling to unravel complex processes shaping spatial patterns of genetic diversity that simultaneously

involve several key environmental variables, notably in landscapes with high spatial co-variation such as river networks.

General conclusion

When adapted to pairwise matrices, causal modeling allows the assessment of complex competing causal models, depicting the *a priori* hypotheses concerning causal relationships among explanatory variables. The proposed framework constitutes a promising alternative to the causal modeling procedure proposed by Cushman et al. (2006), as it allows the assessment of both direct and indirect causal relationships among numerous predictors.

Nevertheless, caution must be taken when using causal modeling. Causal modeling procedures rely directly on the formally stated *a priori* causal hypotheses depicted in the initial causal model. As a first consequence, inferred relationships among variables cannot be considered as absolute causal links; rather, these relationships can only be considered as possible causal links because some important but unknown (or unmeasured) variables may have been overlooked. Although investigation of the interplay between direct and indirect relationships in the optimal model may reveal hidden pathways, thus shedding light on the biological processes acting on the dependent variable, researchers should always keep in mind that a causal model cannot provide information beyond stated *a priori* hypotheses. Our empirical dataset exemplifies this observation appropriately, with genetic drift identified as a possible driver of spatial genetic variation in *G. occitaniae* only under the hypothesis of a direct relationship between oxygen concentration and the effective population size. As a second consequence, causal modeling cannot be confidently considered as a data mining procedure, as investigation of correlation coefficients in the absence of any implicit *a priori* hypothesis (and thus in the absence of any formal causal model) may produce spurious conclusions (Legendre and Legendre 2012; Prunier et al. 2015). Note also that the interpretation of causal modeling becomes more complex as the number of predictors increases. The keys toward the successful use of causal modeling are thus (i) a well thought-out initial set of possible causal models, (ii) a cautious interpretation of the combination of AIC, p-values and confidence intervals, together providing a body of evidence as to the relevance of considered models, and (iii) proper biological interpretation of inferred direct and indirect relationships in the light of formally stated *a priori* hypotheses. Keeping these prerequisites in mind, we advocate the use of causal modeling as a powerful explanatory tool in landscape genetics.

I.9 - References

Aljanabi, S. M., and I. Martinez. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research 25:4692–4693.

Balkenhol, N., L. P. Waits, and R. J. Dezzani. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830.

Belmar-Lucero, S., J. L. A. Wood, S. Scott, A. B. Harbicht, J. A. Hutchings, and D. J. Fraser. 2012. Concurrent habitat and life history influences on effective/census population size ratios in stream-dwelling trout. Ecology and Evolution 2:562–573.

Blanchet, S., O. Rey, R. Etienne, S. Lek, and G. Loot. 2010. Species-specific responses to landscape fragmentation: implications for management strategies. Evolutionary Applications 3:291–304.

Bollen, K. A. 1989. Structural equations with latent variables. Wiley, New York, NY.

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2013. Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. Evolution 67:3258–3273.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd edition.). Springer Verlag, New York, NY.

Campbell Grant, E. H., W. H. Lowe, and W. F. Fagan. 2007. Living in the branches: population dynamics and ecological processes in dendritic networks. Ecology Letters 10:165–175.

Cardon, M., G. Loot, G. Grenouillet, and S. Blanchet. 2011. Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal analysis. Journal of Animal Ecology 80:657–667.

Carrara, F., F. Altermatt, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. Dendritic connectivity controls biodiversity patterns in experimental metacommunities. Proceedings of the National Academy of Sciences 109:5761–5766.

Clarke, R. T., P. Rothery, and A. F. Raybould. 2002. Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. Journal of Agricultural, Biological, and Environmental Statistics 7:361.

Costedoat, C., and A. Gilles. 2009. Quaternary Pattern of Freshwater Fishes in Europe: Comparative Phylogeography and Conservation Perspective. The Open Conservation Biology Journal 3:36–48.

Crispo, E., P. Bentzen, D. N. Reznick, M. T. Kinnison, and A. P. Hendry. 2006. The relative influence of natural selection and geography on gene flow in guppies. Molecular Ecology 15:49–62.

Crispo, E., and L. J. Chapman. 2008. Population genetic structure across dissolved oxygen regimes in an African cichlid fish. Molecular Ecology 17:2134–2148.

Cushman, S. A., and E. L. Landguth. 2010. Spurious correlations and inference in landscape genetics. Molecular Ecology 19:3592–3602.

Cushman, S. A., K. S. McKelvey, J. Hayden, and M. K. Schwartz. 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. The American Naturalist 168:486–499.

Cushman, S. A., T. N. Wasserman, E. L. Landguth, and A. J. Shirk. 2013. Re-evaluating causal modeling with mantel tests in landscape genetics. Diversity 5:51–72.

Dray, S., and A.-B. Dufour. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software 22:1–20.

Ewers, R. M., and R. K. Didham. 2006. Confounding factors in the detection of species responses to habitat fragmentation. Biological Reviews 81:117–142.

Faulks, L. K., D. M. Gilligan, and L. B. Beheregaray. 2011. The role of anthropogenic vs. natural in-stream structures in determining connectivity and genetic diversity in an endangered freshwater fish, Macquarie perch (Macquaria australasica). Evolutionary Applications 4:589–601.

Fisher, R. A. 1938. Statistical Methods for Research Workers (7th edition.). Oliver & Boyd, Edinburgh.

———. 1950. Contributions to mathematical statistics. Wiley, New York, NY.

Fourtune, L., I. Paz-Vinas, G. Loot, J. G. Prunier, and S. Blanchet. 2016. Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. Freshwater Biology 61:1830–1845.

Freeman, L. 1977. A Set of Measures of Centrality Based on Betweenness. Sociometry 40:35–41.

Goodman, S. N. 1999. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. Annals of Internal Medicine 130:995.

Goslee, S. C., and D. L. Urban. 2007. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. Journal of Statistical Software 22:1–19.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices

(version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995).

Grace, J. B. 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press, Cambridge.

Graves, T. A., P. Beier, and J. A. Royle. 2013. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. Molecular Ecology 22:3888–3903.

Guillot, G., R. Leblois, A. Coulon, and A. C. Frantz. 2009. Statistical methods in spatial genetics. Molecular Ecology 18:4734–4756.

Hagberg, A. A., D. A. Schult, and P. J. Swart. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. Pages 11–15 *in* G. Varoquaux, T. Vaught, and J. Millman, eds. Proceedings of the 7th Python in Science Conference (ScyPy2008). Pasadena, CA.

Hutchison, D., and A. Templeton. 1999. Correlation of pairwise genetic and geographic distance measure: inferring the relative influences of gene flow and drift on distribution of genetic variability. Evolution 53:1898–1914.

Jaquiéry, J., T. Broquet, A. H. Hirzel, J. Yearsley, and N. Perrin. 2011. Inferring landscape effects on dispersal from genetic distances: how far can we go? Molecular Ecology 20:692–705.

Jost, L. 2008. GST and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026.

Keller, D., R. Holderegger, M. J. van Strien, and J. Bolliger. 2015. How to make landscape genetics beneficial for conservation management? Conservation Genetics 16:503–512.

Landguth, E., S. A. Cushman, and N. Balkenhol. 2015. Simulation Modeling in Landscape Genetics. Pages 99–113 *in* N. Balkenhol, S. A. Cushman, A. T. Storfer, and L. P. Waits, eds. Landscape Genetics: Concepts, Methods, Applications. Wiley-Blackwell, Oxford.

Legendre, P. 2000. Comparison of permutation methods for the partial correlation and partial mantel tests. Journal of Statistical Computation and Simulation 67:37–73.

Legendre, P., and M.-J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources 10:831–844.

Legendre, P., M.-J. Fortin, and D. Borcard. 2015. Should the Mantel test be used in spatial analysis? Methods in Ecology and Evolution.

Legendre, P., and L. F. J. Legendre. 2012. Numerical Ecology. Elsevier, Amsterdam.

Lichstein, J. W. 2007. Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecology 188:117–131.

Manel, S., and R. Holderegger. 2013. Ten years of landscape genetics. Trends in Ecology & Evolution 28:614–621.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189–197.

Marchetti, G. M. 2006. Independencies Induced from a Graphical Markov Model After Marginalization and Conditioning: The R Package ggm. Journal of Statistical Software 15:1–15.

Oberski, D. L. 2014. lavaan. survey: An R Package for Complex Survey Analysis of Structural Equation Models. Journal of Statistical Software 57:1–27.

ONEMA. 2010. Référentiel national des Obstacles à l'Ecoulement : une cartographie nationale des obstacles sur les cours d'eau. Les Fiches de l'ONEMA.

Paz-Vinas, I., and S. Blanchet. 2015. Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. Journal of Evolutionary Biology 28:986–994.

Paz-Vinas, I., G. Loot, V. M. Stevens, and S. Blanchet. 2015. Evolutionary processes driving spatial patterns of intra-specific genetic diversity in river ecosystems. Molecular Ecology 24:4586–4604.

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA.

———. 2009. Causality: models, reasoning, and inference (Second Edition.). Cambridge University Press, Cambridge ; New York, NY.

Pearl, J., and T. Verma. 1987. The logic of representing dependencies by directed graphs. Pages 374–379 *in* Proceedings of the sixth National conference on Artificial intelligence - Volume 1, AAAI'87. AAAI Press, Seattle, WA.

Pella, H., J. Lejot, N. Lamouroux, and T. Snelder. 2012. Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. Géomorphologie : relief, processus, environnement 3:317–336.

Prunier, J. G., M. Colyn, X. Legendre, K. F. Nimon, and M. C. Flamand. 2015. Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. Molecular Ecology 24:263–283.

Quantum GIS Development Team. 2017. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project.

R Development Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raeymaekers, J. A. M., D. Raeymaekers, I. Koizumi, S. Geldof, and F. A. M. Volckaert. 2009. Guidelines for restoring connectivity around water mills: a population genetic approach to the management of riverine fish. Journal of Applied Ecology 46:562–571.

Rosseel, Y. 2012. Lavaan: an R package for structural equation modeling. Journal of Statistical Software 48:1–36.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336–352.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution 1:103–113.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann. 2014. Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? Evolution 68:1–15.

Shipley, B. 2000*a*. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge.

———. 2000*b*. A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling 7:206–218.

———. 2003. Testing Recursive Path Models With Correlated Errors Using D-Separation. Structural Equation Modeling 10:214–221.

———. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.

———. 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology 94:560–564.

Smith, T. B., and L. Bernatchez. 2008. Evolutionary change in human-altered environments. Molecular Ecology 17:1–8.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. Systematic Zoology 35:627–632.

Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits. 2010. Landscape genetics: where are we now? Molecular Ecology 19:3496–3514.

Van Strien, M. J., D. Keller, and R. Holderegger. 2012. A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. Molecular Ecology 21:4010–4023.

Winter, D. J. 2012. MMOD: an R library for the calculation of population differentiation statistics. Molecular ecology resources 12:1158–1160.

Wright, S. 1921. Correlation and Causation. Journal of agricultural research 20:557–585.

I.10 - Acknowledgments

We are very grateful to all the ONEMA staff that contributed to the field sampling, in particular Laurence Blanc. Leslie Faggiano, Charlotte Evangelista, Christine Lauzeral, Loïc Tudesque, Roselyne Etienne, Sébastien Villéger, Julien Cucherousset, Vincent Dubut, Alain Blanchet and South-Western France ONEMA staff are thanked for their help in the field. Charlotte Veyssière, Roselyne Etienne and Vincent Dubut are also thanked for their help with molecular analyses. We warmly thanks Dr. Andrew Hendry for insightful comments on the paper. The authors also thank the "Génopole Toulouse" for help with genotyping. LF was financially supported by a MESR ("Ministère de l'Enseignement Supérieur et de la Recherche") PhD scholarship during this study. This study was part of the European project “IMPACT”. This project has been carried out with financial support from the Commission of the European Communities, specific RTD programme "IWRMNET". This work has been done in two research units (EDB & SETE) that are part of the "Laboratoire d’Excellence" (LABEX) entitled TULIP (ANR-10-LABX-41).

I.11 Supplementary materials for Chapter I

Appendix I-S1: User-friendly R functions for the application of the permutation-based d-sep test, clustering-based path analysis, permutation-based path analysis and a parametric bootstrap procedure dedicated to the handling of pairwise matrices in path analysis. The permutation-based d-sep test and clustering-based path analysis assess the relative and absolute fit of causal models, clustering-based path analysis and permutation-based path analysis provide corrected p-values for path coefficients and the parametric bootstrap procedure computes confidence intervals for the path coefficients.

Permutation-based d-sep test

Permutation-based d-sep test is applied through the dsep.test function. It provides as output the C-value, number of degrees of freedom, p-value and AIC of the tested model (Shipley

2009). It takes as arguments:

- model: a model written in the lavaan format of ‘lavaan’ package (Rosseel 2012);

- data: a data set which must be a list of pairwise matrices to use MRM;

- method: the method used to test d-separations. Must be set to “MRM” when applied to pairwise data. “lm” can be used as an alternative for non-pairwise data;

- nperm: number of permutations to be performed by *MRM* function of ‘ecodist’ package (Goslee and Urban 2007); default is 1000.

```
dsep.test = function(model, data, method, nperm = 1000, nboot = 1000)
{
 require(ggm) ; require(ecodist) ; require(stringr)

 data_vec = NULL
 if(class(data) == "list"){
  npop = dim(data[[1]])[1]
  nvar = length(data)
  for(i in 1:nvar){
   temp = as.vector(as.dist(data[[i]]))
   data_vec = cbind(data_vec, temp)
  }
  colnames(data_vec) = names(data)
  npairs = (npop/2*(npop-1))
  sites = as.factor(seq(1, npop, 1))
  data_vec = as.data.frame(data_vec)
 }else{
  data_vec = as.data.frame(data)
  npop = dim(data)[1]
 }
 createDAG = function(model)
 {
  if (model == ''){
   amat = matrix(0, length(data), length(data)) ; colnames(amat) = names(data) ;
rownames(amat) = names(data)
  } else {
   model = str_replace_all(str_trim(model), '\n', ',')
   f = as.list(as.vector(strsplit(model, ",")[[1]]))
   for (i in c(1:length(f))){
    f[[i]]=as.formula(f[[i]])
   }
   nb = length(f)
   nod = c()
   for (k in 1:nb) {
    tt = terms(f[[k]], specials = "I")
    vars = dimnames(attr(tt, "factors"))[[1]]
    skip = attr(tt, "specials")$I
    if (!is.null(skip))
```

```
      vars = vars[-skip]
    nod = c(nod, vars)
   }
   N = unique(nod)
   dN = length(N)
   amat = matrix(0, dN, dN)
   for (k in 1:nb) {
    tt = terms(f[[k]], specials = "I")
    vars = dimnames(attr(tt, "factors"))[[1]]
    if (attr(tt, "response") == 1) {
     j = match(vars[1], N)
     i = match(vars[-1], N)
     amat[i, j] = 1
    }
    else if (attr(tt, "response") == 0)
     stop("Some equations have no response")
   }
   if (!isAcyclic(amat))
    warning("The graph contains directed cycles!")
   dimnames(amat) = list(N, N)
  }
  amat
}

createREG = function(model)
{
 model = str_replace_all(str_trim(model), '\n', ',')
 f = as.list(as.vector(strsplit(model, ",")[[1]]))
 for (i in c(1:length(f))){
  f[[i]]=as.formula(f[[i]])
 }
 f
}

model_DAG = createDAG(model)
for (i in 1:length(data)){
 if ((names(data)[i] %in% colnames(model_DAG)) == FALSE){
  model_DAG = cbind(model_DAG, rep(0, dim(model_DAG)[1]))
  colnames(model_DAG)[dim(model_DAG)[2]] = names(data)[i]
  model_DAG = rbind(model_DAG, rep(0, dim(model_DAG)[2]))
  rownames(model_DAG)[dim(model_DAG)[1]] = names(data)[i]
 }
}

model_REG = createREG(model)

basis = basiSet(model_DAG)
pval = rep(-1, length(basis))
for (i in 1:length(basis))
```

```
 {
  cond = length(basis[[i]]) - 2
  dsep = paste(basis[[i]][1], '~', basis[[i]][2])
  if (cond > 0)
  {
   for (j in 1:cond)
   {
    dsep = paste(dsep, '+', basis[[i]][2+j])
   }
  }
  if (method == 'lm')
  {
   lm = lm(dsep, data = data_vec)
   pval[i] = summary(lm)[[4]][2,4]
  } else
  {
   if (method == 'MRM')
   {
    mrm = MRM(dsep, nperm = nperm, data = data_vec)
    pval[i] = mrm$coef[2,2]
   } else
   {
    stop('This method is not available. Use \'lm\' or \'MRM\' instead.')
   }
  }
 }
 ctest = -2*sum(log(pval))
 df = 2 * length(basis)
 pvalue = 1 - pchisq(ctest, df)
 q = 0
 for (i in 1:length(model_REG)){
  reg = as.formula(model_REG[[i]])
  if(method == "lm"){
   fit = lm(reg, data = data_vec)
   q = q + length(coef(fit)) + 1
  } else {
   if(method == "MRM"){
    fit = MRM(reg, data = data_vec)
    q = q + length(fit$coef[,1]) + 1
   }
  }
 }
 AIC = ctest + 2*q
 output = list(Ctest = ctest, df = df, Pvalue = pvalue, AIC = AIC)
 return(output)
}
```

Clustering-based path analysis

Clustering-based path analysis is applied through the clustering.based.pathanalysis function. It provides as output the F_{ML} statistics, number of degree of freedom, p-value and AIC (following the formula given by Grace 2006) of a given model, as well as the values and associated p-values of its path coefficients. It takes as arguments:

- model: a model written in the lavaan format of 'lavaan' package (Rosseel 2012);

- data: a data set which can be a list of pairwise matrices or a data frame containing the variables in the model as columns of dissimilarities and in which two columns gather the identities of the two sites implied in each pair;

- Id1: the name of the column gathering the identities of the first site (only necessary when data is a data frame), default is "ID1";

- Id2: the name of the column gathering the identities of the second site (only necessary when data is a data frame), default is "ID2";

- nperm: number of permutations to perform, default is 1000;

- verb: if set to TRUE, the index value is printed at each iteration, default is FALSE.

```
clustering.based.pathanalysis = function(model, data, Id1 = "ID1", Id2 = "ID2", nperm =
1000, verb = F)
{
 require(survey) ; require(lavaan) ; require(lavaan.survey)
 data_vec = NULL
 if(class(data) == "list"){
  npop = dim(data[[1]])[1]
  nvar = length(data)
  for(i in 1:nvar){
   temp = as.vector(as.dist(data[[i]]))
   data_vec = cbind(data_vec, temp)
  }
  colnames(data_vec) = names(data)
  npairs = (npop/2*(npop-1))
  sites = as.factor(seq(1, npop, 1))
  ID1 = NULL ; ID2 = NULL
  for (i in 1:(npop-1)) {
   ID2 = c(ID2, rep(sites[i], npop - i))
   ID1 = c(ID1, sites[(i+1):npop])
  }
  data_vec = as.data.frame(cbind(data_vec, ID1, ID2))
  data_vec$ID1 = as.factor(data_vec$ID1)
  data_vec$ID2 = as.factor(data_vec$ID2)
 }else{
  data_vec = as.data.frame(data)
 }
```

```
  fit = do.call("sem",list(model = model, data = data_vec, warn = FALSE))

  colID1 = which(names(data_vec) == Id1)
  colID2 = which(names(data_vec) == Id2)

  data_perm = data_vec
  names(data_perm)[colID1] = "ID1"
  names(data_perm)[colID2] = "ID2"
  levels(data_perm[, colID1]) = c(levels(data_perm[, colID1]),  levels(data_perm[, colID2]))
  levels(data_perm[, colID2]) = c(levels(data_perm[, colID1]),  levels(data_perm[, colID2]))

  pval_model = NULL
  pval_coeff = NULL
  chisq = NULL

  for (i in 1:nperm){
    for (j in c(1:dim(data_perm)[1])){
      rand = sample(0:1,1)
      if (rand == 1){
        idd1 = data_perm[j, colID1]
        idd2 = data_perm[j, colID2]
        data_perm[j, colID1] = idd2
        data_perm[j, colID2] = idd1
      }
    }
    survey_design = svydesign(ids = ~ ID1 + ID2, prob = ~1, data = data_perm)
    fit_survey = lavaan.survey(fit, survey_design)
    pval_model = c(pval_model, fit_survey@Fit@test[[2]]$pvalue)
    pval_coeff = cbind(pval_coeff,
standardizedSolution(fit_survey)[standardizedSolution(fit_survey)$op == '~',]$pvalue)
    chisq = cbind(chisq, fit_survey@Fit@test[[2]]$stat)
    if(verb == T){print(i)}
  }
  results = list(Model = NULL, Coefficients = NULL)
  results$Model = mean(pval_model)
  for (i in c(1:dim(pval_coeff)[1])){
    results$Coefficients = rbind(results$Coefficients, mean(pval_coeff[i,]))
  }
  colnames(results[[2]]) = names(results[[1]])
  FML = mean(chisq)
  df = fit_survey@Fit@test[[2]]$df
  AIC = mean(chisq) - 2*df
  output = list(Model = NULL, Coefficients = NULL)
  output$Model = c(FML, df, results$Model, AIC)
  names(output$Model) = c('FML', 'df', 'Pvalue', 'AIC')
  coeffs_class = standardizedSolution(fit)[standardizedSolution(fit)$op == '~',]
  coeffs_class[,1] = paste(coeffs_class[,1], coeffs_class[,2], coeffs_class[,3])
  coeffs_class = coeffs_class[,-c(2,3,5:7)]
```

```
  row.names(coeffs_class) = NULL
  output$Coefficients = cbind(coeffs_class, results$Coefficients)
  names(output$Coefficients) = c('Path', 'Standardized estimate', 'Pvalue')
  return(output)
}
```

Permutation-based path analysis

Permutation-based path analysis is applied through the permutation.based.pathanalysis function. It provides as output the values and associated p-values of the path coefficients of a given model. It takes as arguments:

- model: a model written in the lavaan format of 'lavaan' package (Rosseel 2012);
- data: a data set which must be a list of pairwise matrices;
- nperm: number of permutations to perform, default is 1000;
- verb: if set to TRUE, the index value is printed at each iteration, default is FALSE.

```
permutation.based.pathanalysis = function(model, data, nperm = 1000, verb = F)
{
  require(MASS) ; require(lavaan)

  npop = dim(data[[1]])[1]
  nvar = length(data)
  data_vec = NULL

  for(i in 1:nvar){
    temp = as.vector(as.dist(data[[i]]))
    data_vec = cbind(data_vec, temp)
  }
  colnames(data_vec) = names(data)

  fit = sem(model, data = data_vec, warn = F)
  coeff = standardizedSolution(fit)$est.std[standardizedSolution(fit)$op == '~']

  coeff_perm = NULL
  pval_coef_perm = NULL

  data_perm = data_vec
  for (i in 1:nperm){
    check_conv = 0
    while (check_conv == 0){
      for (j in 1:nvar){
        rarray = sample(npop)
        data_perm[,j] = as.vector(as.dist(data_list[[j]][rarray,rarray]))
      }
      fit_perm = sem(model, data = data_perm, warn = F)
```

```
    check_conv = round(sum(inspect(fit_perm, 'rsquare')), digit = 3)
   }
   coeff_perm = cbind(coeff_perm,
standardizedSolution(fit_perm)$est.std[standardizedSolution(fit_perm)$op == '~'])
   if(verb == T){print(i)}
  }

  for (i in 1:length(coeff)){
   if (coeff[i]>=0){
    pval_coef_perm = c(pval_coef_perm, sum(coeff_perm[i,]>coeff[i])/(nperm + 1))
   } else {
    pval_coef_perm = c(pval_coef_perm, sum(coeff_perm[i,]<coeff[i])/(nperm + 1))
   }
  }
  coeffs_class = standardizedSolution(fit)[standardizedSolution(fit)$op == '~',]
  coeffs_class[,1] = paste(coeffs_class[,1], coeffs_class[,2], coeffs_class[,3])
  coeffs_class = coeffs_class[,-c(2,3,5:7)]
  coeffs = cbind(coeffs_class, pval_coef_perm)
  names(coeffs) = c('Path', 'Standardized estimate', 'Pvalue')
  return(coeffs)
}
```

A bootstrap procedure dedicated to the handling of pairwise matrices in a path analysis framework.

This procedure allows the computation of confidence intervals for the path coefficients of a causal model. It is applied through the pairwise.bootstrap.pa function. It provides as output the values (estimated through the sem function of 'lavaan' package; Rosseel 2012) and associated 95% confidence intervals of the path coefficients of a given model. It takes as arguments:

- model: a model written in the lavaan format of 'lavaan' package (Rosseel 2012);
- data: a data set which must be a list of pairwise matrices;
- nboot: number of bootstrap replicates, default is 1000;

```
pairwise.bootstrap.pa = function(model, data, nboot = 1000){
  require(lavaan)
  prep = function(x){ x=data.matrix(x)
  x=x[upper.tri(x, diag = FALSE)]}
  mylist2=lapply(mylist, prep)
  mydata=as.data.frame(mylist2)
  mysem = sem(model, mydata)
  mysemresult=standardizedSolution(mysem)[standardizedSolution(mysem)$op == '~',]
  res_pa_coef = as.data.frame(matrix(NA, nrow = dim(mysemresult)[1], ncol = 4))
```

```
 res_pa_coef[,1]=paste(mysemresult[,1], mysemresult[,2], mysemresult[,3])
 res_pa_coef[,2]=mysemresult$est.std
 colnames(res_pa_coef)=c('Path', 'Standardized estimate','lowCI','uppCI')

resboot=as.data.frame(matrix(NA,ncol=nboot,nrow=dim(standardizedSolution(mysem)[stand
ardizedSolution(mysem)$op == '~',])[1]))
 for (j in c(1:nboot)){
  #print(j)
  mybootdata=mydata
  colnames(mybootdata)=names(mylist)
  rarray=sort(sample(dim(mylist[[1]])[1],dim(mylist[[1]])[1],replace=T))
  for (k in 1:length(mylist)){ mybootdata[,k]=prep(mylist[[k]][rarray,rarray])}
  fitboot = sem(model, mybootdata)
  resboot[,j]=standardizedSolution(fitboot)[standardizedSolution(fitboot)$op == '~',4]
 }
 res_pa_coef[,3:4]=t(apply(resboot,1,function(x) quantile(x,c(.025,.975))))
 return(res_pa_coef)
}
```

References for Appendix I-S1:

Goslee, S. C., and D. L. Urban. 2007. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. Journal of Statistical Software 22:1–19.

Grace, J. B. 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press, Cambridge.

Rosseel, Y. 2012. Lavaan: an R package for structural equation modeling. Journal of Statistical Software 48:1–36.

Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.

Appendix I-S2: Scripts used for the simulation test.

```
#####Computing “inadequate”, “intermediate” and “adequate” models AIC and p-values with path analysis, clustering-based path analysis, the d-sep test and the permutation-based d-sep test#####
pval_pa_classic = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(pval_pa_classic) = c("model1", "model2", "model3")
AIC_pa_classic = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(AIC_pa_classic) = c("model1", "model2", "model3")
pval_dsep_classic = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(pval_dsep_classic) = c("model1", "model2", "model3")
AIC_dsep_classic = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(AIC_dsep_classic) = c("model1", "model2", "model3")
```

```
pval_pa_new = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(pval_pa_new) = c("model1", "model2", "model3")
AIC_pa_new = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(AIC_pa_new) = c("model1", "model2", "model3")
pval_dsep_new = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(pval_dsep_new) = c("model1", "model2", "model3")
AIC_dsep_new = matrix(rep(NA, 3000), nrow = 1000, ncol = 3)
colnames(AIC_dsep_new) = c("model1", "model2", "model3")

# “npop” is the number of sites in the simulated data set and “sites” assigns a number from 1 to 100 to each of them
npop = 50
npairs = (npop/2*(npop-1))
sites = as.factor(seq(1, npop, 1))
#ID1 and ID2 are the identities of the two sites implied in each pairwise value of the simulated data set
ID1 = NULL ; ID2 = NULL
for (i in 1:(npop-1)) {
  ID2 = c(ID2, rep(sites[i], npop - i))
  ID1 = c(ID1, sites[(i+1):npop])
}

set.seed(42)
for (i in 1:100){
  #Creating variables
  x1 = rnorm(npop, mean = 4, sd = 1)
  x2 = rnorm(npop, mean = 5, sd = 1)
  x1_dist = dist(x1)
  x2_dist = dist(x2)

  x3 = 5 + runif(1, 0.8, 1.6)*x1 + rnorm(npop, 0, sd = 2)
  x4 = 3 + runif(1, 0.8, 1.6)*x2 + runif(1, 0.8, 1.6)*x3 + rnorm(npop, 0, sd = 2)
  x5 = 4 + runif(1, 0.8, 1.6)*x2  + rnorm(npop, 0, sd = 2)
  x3_dist = dist(x3)
  x4_dist = dist(x4)
  x5_dist = dist(x5)

  #Scaling and creating data object
  x1_dist = (x1_dist-mean(x1_dist))/sd(x1_dist)
  x2_dist = (x2_dist-mean(x2_dist))/sd(x2_dist)
  x3_dist = (x3_dist-mean(x3_dist))/sd(x3_dist)
  x4_dist = (x4_dist-mean(x4_dist))/sd(x4_dist)
  x5_dist = (x5_dist-mean(x5_dist))/sd(x5_dist)

  data_vec = as.data.frame(cbind(as.vector(x1_dist), as.vector(x2_dist), as.vector(x3_dist), as.vector(x4_dist), as.vector(x5_dist), ID1, ID2))
  colnames(data_vec) = c('x1_dist', 'x2_dist', 'x3_dist', 'x4_dist', 'x5_dist', 'ID1', 'ID2')
  data_vec$ID1 = as.factor(data_vec$ID1)
```

```
data_vec$ID2 = as.factor(data_vec$ID2)
data_list = list(x1_dist = as.matrix(x1_dist), x2_dist = as.matrix(x2_dist), x3_dist = as.matrix(x3_dist), x4_dist = as.matrix(x4_dist), x5_dist = as.matrix(x5_dist))

#Adequate model (figure 1A)
model1 = '
x3_dist ~ x1_dist
x4_dist ~ x3_dist + x2_dist
x5_dist ~ x2_dist
'
#Intermediate model (figure 1A)
model2 = '
x3_dist ~ x1_dist + x2_dist
x4_dist ~ x1_dist
x5_dist ~ x2_dist
'
#Inadequate model (figure 1A)
model3 = '
x4_dist ~ x1_dist + x5_dist
x3_dist ~ x2_dist
x5_dist ~ x1_dist
'

fit_1 = sem(model1, data = data_vec, warn = F)
fit_2 = sem(model2, data = data_vec, warn = F)
fit_3 = sem(model3, data = data_vec, warn = F)
fit_1clust = clustering.based.pathanalysis(model1, data_vec, nperm = 1000)
fit_2clust = clustering.based.pathanalysis(model2, data_vec, nperm = 1000)
fit_3clust = clustering.based.pathanalysis(model3, data_vec, nperm = 1000)
fit_dseptestlm_1 = dsep.test(model1, data_list, method = 'lm')
fit_dseptestlm_2 = dsep.test(model2, data_list, method = 'lm')
fit_dseptestlm_3 = dsep.test(model3, data_list, method = 'lm')
fit_dseptestMRM_1 = dsep.test(model1, data_list, method = 'MRM', 1000)
fit_dseptestMRM_2 = dsep.test(model2, data_list, method = 'MRM', 1000)
fit_dseptestMRM_3 = dsep.test(model3, data_list, method = 'MRM', 1000)

pval_pa_classic[i,1] = inspect(fit_1, "fit")[5]
pval_pa_classic[i,2] = inspect(fit_2, "fit")[5]
pval_pa_classic[i,3] = inspect(fit_3, "fit")[5]

pval_pa_new[i,1] = fit_1clust$Model[3]
pval_pa_new[i,2] = fit_2clust$Model[3]
pval_pa_new[i,3] = fit_3clust$Model[3]

AIC_pa_classic[i,1] = inspect(fit_1, "fit")[3] - 2*inspect(fit_1, "fit")[4]
AIC_pa_classic[i,2] = inspect(fit_2, "fit")[3] - 2*inspect(fit_2, "fit")[4]
AIC_pa_classic[i,3] = inspect(fit_3, "fit")[3] - 2*inspect(fit_3, "fit")[4]

AIC_pa_new[i,1] = fit_1clust$Model[4]
```

```
  AIC_pa_new[i,2] = fit_2clust$Model[4]
  AIC_pa_new[i,3] = fit_3clust$Model[4]

  pval_dsep_classic[i,1] = fit_dseptestlm_1$Pvalue
  pval_dsep_classic[i,2] = fit_dseptestlm_2$Pvalue
  pval_dsep_classic[i,3] = fit_dseptestlm_3$Pvalue

  pval_dsep_new[i,1] = fit_dseptestMRM_1$Pvalue
  pval_dsep_new[i,2] = fit_dseptestMRM_2$Pvalue
  pval_dsep_new[i,3] = fit_dseptestMRM_3$Pvalue

  AIC_dsep_classic[i,1] = fit_dseptestlm_1$AIC
  AIC_dsep_classic[i,2] = fit_dseptestlm_2$AIC
  AIC_dsep_classic[i,3] = fit_dseptestlm_3$AIC

  AIC_dsep_new[i,1] = fit_dseptestMRM_1$AIC
  AIC_dsep_new[i,2] = fit_dseptestMRM_2$AIC
  AIC_dsep_new[i,3] = fit_dseptestMRM_3$AIC

}
```

#####Computation of path coefficients p-values and CI using classical path analysis, clustering-based path analysis, permutation-based path analysis from simulated data sets and the parametric bootstrap procedure dedicated to the handling of pairwise matrices#####

```
# “Adequate coefficients” are in columns 1, 4, 6 and 7 and “inadequate coefficients” are in columns 2, 3, 5 and 8 (figure 2A).
res_pa_coef = matrix(NA, 1000, 6)
colnames(res_pa_coef) = c('x3 ~ x1', 'x3 ~ x2', 'x5 ~ x1', 'x5 ~ x2', 'x4 ~ x5', 'x4 ~ x3')
res_pa_coef_clust = matrix(NA, 1000, 6)
colnames(res_pa_coef_clust) = c('x3 ~ x1', 'x3 ~ x2', 'x5 ~ x1', 'x5 ~ x2', 'x4 ~ x5', 'x4 ~ x3')
res_pa_coef_permut = matrix(NA, 1000, 6)
colnames(res_pa_coef_permut) = c('x3 ~ x1', 'x3 ~ x2', 'x5 ~ x1', 'x5 ~ x2', 'x4 ~ x5', 'x4 ~ x3')
res_pa_coef_CI_low = matrix(NA, 1000, 6)
colnames(res_pa_coef_CI_low) = c('x3 ~ x1', 'x3 ~ x2', 'x5 ~ x1', 'x5 ~ x2', 'x4 ~ x5', 'x4 ~ x3')
res_pa_coef_CI_up = matrix(NA, 1000, 6)
colnames(res_pa_coef_CI_up) = c('x3 ~ x1', 'x3 ~ x2', 'x5 ~ x1', 'x5 ~ x2', 'x4 ~ x5', 'x4 ~ x3')

npop = 50
npairs = (npop/2*(npop-1))
sites = as.factor(seq(1, npop, 1))
ID1 = NULL ; ID2 = NULL
for (i in 1:(npop-1)) {
  ID2 = c(ID2, rep(sites[i], npop - i))
  ID1 = c(ID1, sites[(i+1):npop])
}
set.seed(42)
for (i in 1:1000){
  #Creating variables
  x1 = rnorm(npop, mean = 4, sd = 1)
```

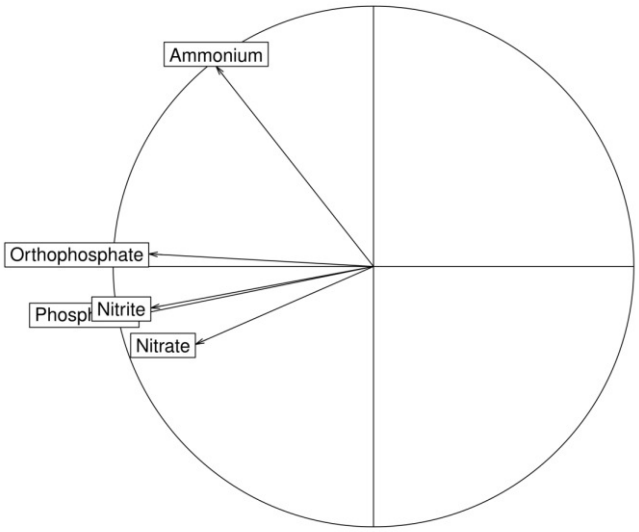
```
  x2 = rnorm(npop, mean = 5, sd = 1)
  x1_dist = dist(x1)
  x2_dist = dist(x2)

  x3 = 5 + runif(1, 0.8, 1.6)*x1 + rnorm(npop, 0, sd = 2)
  x4 = 3 + runif(1, 0.8, 1.6)*x2 + runif(1, 0.8, 1.6)*x3 + rnorm(npop, 0, sd = 2)
  x5 = 4 + runif(1, 0.8, 1.6)*x2  + rnorm(npop, 0, sd = 2)
  x3_dist = dist(x3)
  x4_dist = dist(x4)
  x5_dist = dist(x5)

  #Scaling and creating data object
  x1_dist = (x1_dist-mean(x1_dist))/sd(x1_dist)
  x2_dist = (x2_dist-mean(x2_dist))/sd(x2_dist)
  x3_dist = (x3_dist-mean(x3_dist))/sd(x3_dist)
  x4_dist = (x4_dist-mean(x4_dist))/sd(x4_dist)
  x5_dist = (x5_dist-mean(x5_dist))/sd(x5_dist)

  data_vec = as.data.frame(cbind(as.vector(x1_dist), as.vector(x2_dist), as.vector(x3_dist),
                           as.vector(x4_dist), as.vector(x5_dist), ID1, ID2))
  colnames(data_vec) = c('x1_dist', 'x2_dist', 'x3_dist', 'x4_dist', 'x5_dist', 'ID1', 'ID2')
  data_vec$ID1 = as.factor(data_vec$ID1)
  data_vec$ID2 = as.factor(data_vec$ID2)
  data_list = list(x1_dist = as.matrix(x1_dist), x2_dist = as.matrix(x2_dist), x3_dist =
as.matrix(x3_dist),
                x4_dist = as.matrix(x4_dist), x5_dist = as.matrix(x5_dist))

  model = '
  x3_dist ~ x1_dist + x2_dist
  x5_dist ~ x1_dist + x2_dist
  x4_dist ~ x5_dist + x3_dist
  '


  fit = sem(model, data_vec)
  fit_clust = clustering.based.pathanalysis(model, data_list, nperm = 1000)$Coefficients
  fit_permut = permutation.based.pathanalysis(model, data_list, nperm = 1000)
  fit_CI = pairwise.bootstrap.pa(model, data_list, nboot = 1000)

  res_pa_coef[i,] = standardizedSolution(fit)$pvalue[standardizedSolution(fit)$op == '~']
  res_pa_coef_clust[i,] = fit_clust$Pvalue
  res_pa_coef_permut[i,] = fit_permut$Pvalue
  res_pa_coef_CI_low[i,] = fit_CI[,3]
  res_pa_coef_CI_up[i,] = fit_CI[,4]
}
```

Appendix I-S3: Table gathering the number of individuals and name of the sampled river of each sampling site.

Site	Number of individuals	River
AGOBez	13	L'Agout
AGOLav	25	L'Agout
AGOSal	11	L'Agout
ARIVen	25	L'Ariège
ARIVer	25	L'Ariège
ARZMas	25	L'Arize
AUVGen	25	L'Auvézère
AVEDru	25	L'Aveyron
AVEFen	25	L'Aveyron
AVEGai	25	L'Aveyron
AVEMon	25	L'Aveyron
BAIBro	15	La Baise
BARMon	25	La Petite Barguelonne
CELAmi	25	Le Célé
CELBag	25	Le Célé
CELBou	25	Le Célé
CELClam	29	Le Célé
CELClav	30	Le Célé
CELEul	25	Le Célé
CELMei	25	Le Célé
CERBre	25	La Cère
CIRPre	23	Le Ciron
CORAng	25	La Corrèze
DADMon	25	Le Dadou
DADRéa	22	Le Dadou
DORBri	25	La Dordogne
DORCen	25	La Dordogne
DORCou	25	La Dordogne
DORFle	25	La Dordogne
DORMey	25	La Dordogne
DORPru	25	La Dordogne
DOUMon	25	Le Dourdou
DROCav	25	Le Dropt
DROCou	25	La Dronne
DROFro	25	La Dronne
DROVal	25	La Dronne
EAUMou	25	L'Eau Blanche
ELLTer	25	L'Elle
GARAge	25	La Garonne
GARBou	25	La Garonne
GARCla	20	La Garonne

GARCou	25	La Garonne
GARGag	25	La Garonne
GARJul	25	La Garonne
GARMur	25	La Garonne
GERAur	25	Le Gers
GERChe	15	Le Gers
GERPre	25	Le Gers
HERBes	25	Le Grand Hers
HERCal	25	Le Grand Hers
ISLTre	25	L'Isle
LOTBal	25	Le Lot
LOTBou	25	Le Lot
LOTCah	25	Le Lot
LOTCla	16	Le Lot
LOTEnt	25	Le Lot
LOTLau	25	Le Lot
LOTLiv	25	Le Lot
LOUFou	24	La Louge
OSSMon	25	L'Osse
OSSMou	25	L'Osse
SALGir	20	Le Salat
SALTou	25	Le Salat
SAVEsp	25	La Save
SAVLev	25	La Save
SEGBar	15	Le Ségur
SEGMar	25	Le Ségur
TARAva	25	Le Tarn
TARBro	25	Le Tarn
TARFor	25	Le Tarn
TARMil	25	Le Tarn
TARRab	25	Le Tarn
TARVil	25	Le Tarn
TOULam	24	Le Touch
TRULou	16	La Truyère
VERCah	25	La Vère
VEZLeo	25	La Vézère
VIABan	25	Le Viaur
VIACal	25	Le Viaur
VIACap	25	Le Viaur
VIAJus	25	Le Viaur
VIANav	25	Le Viaur
VIASer	25	Le Viaur

Appendix I-S4: Correlation circle of the PCA performed on five chemical components.



Chapter II - Lessons from the Fishes: A multi-species analysis reveals common processes underlying similar species-genetic diversity correlations

By Lisa Fourtune[1], Ivan Paz-Vinas[2,3], Géraldine Loot[1,4], Jérôme G. Prunier[1], Simon Blanchet[1,3]

[1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS); UMR 5321 (Station d'Écologie Théorique et Expérimentale, SETE), 09200 Moulis, France
[2] Aix-Marseille Université, CNRS, Institut de Recherche pour le Développement (IRD), Avignon Université; UMR 7263 (Institut Méditerranéen de la Biodiversité et d'Ecologie marine et continentale, IMBE), Centre Saint-Charles, Case 36, Marseille, France
[3] CNRS, UPS, École Nationale de Formation Agronomique (ENFA); UMR 5174 (Laboratoire Évolution & Diversité Biologique, EDB), 31062 Toulouse cedex 4, France
[4] Université de Toulouse, UPS; UMR 5174 (EDB), 31062 Toulouse cedex 4, France

Published in **Freshwater Biology**, vol. 61: 1830–1845 (2016)

II.1 - Résumé

Les corrélations entre diversité spécifique et génétique (SGDC) ont été étudiées chez une grande variété d'organismes, ce qui a permis d'améliorer notre compréhension des processus pouvant influencer ces deux niveaux de diversité de manière semblable. Cependant, peu d'études ont comparé les SGDC (ainsi que les processus qui les soutendent) chez plusieurs espèces échantillonnées sur un même paysage.

Dans la présente étude portant sur quatre espèces de poissons d'eau douce (*Barbatula barbatula, Gobio occitaniae, Phoxinus phoxinus* et *Squalius cephalus*) échantillonnées dans un important bassin versant (le bassin versant de Garonne-Dordogne en France), nous avons combiné une approche multi-spécifique à des analyses causales afin de (i) estimer et comparer les SGDC de ces quatre espèces au niveau α et β, et (ii) inférer les processus à l'origine de ces SGDC. La diversité génétique de chaque espèce a été estimée à partir de marqueurs microsatellites. La diversité spécifique a été estimée à partir de la richesse spécifique mesurée lors de pêches électriques. Les conditions environnementales ont été décrites pour les 81 sites d'échantillonnage.

Nous avons obtenu des valeurs de α-SGDC significativement positives pour les quatre espèces d'intérêt et des valeurs de β-SGDC plus faibles et significativement positives chez deux espèces sur quatre seulement. Les analyses causales ont permis d'identifier deux variables (l'isolation géographique et la taille de l'habitat disponible) à l'origine des α-SGDC. Nous avons constaté que, bien que faibles, les β-SGDC étaient dûes à une relation directe entre différentiations génétique et taxonomique et à une influence de l'environnement abiotique agissant comme un filtre sur les espèces et les allèles.

Notre étude montre que des processus écologiques et évolutifs en lien avec le filtrage environnemental, la migration, la dérive et l'historique de colonisation permettent de comprendre simultanément la diversité spécifique et génétique de poissons d'eau douce.

II.2 - Abstract

Species-genetic diversity correlations (SGDCs) have been investigated over a large spectra of organisms, which has greatly improved our understanding of parallel processes potentially driving both species and genetic diversity. However, there are still few studies comparing SGDCs (and underlying processes) for multiple species sampled over a single landscape.

Here, focusing on freshwater fish sampled across a large river basin (the Garonne-Dordogne river basin, France), we combined a multi-species approach and causal analyses to (i) assess and compare both α-SGDCs and β-SGDCs among species, and (ii) infer processes underlying α-SGDCs and β-SGDCs. Genetic, intraspecific, diversity was assessed for four sympatric fish (*Barbatula barbatula, Gobio occitaniae, Phoxinus phoxinus* and *Squalius cephalus*) using microsatellite markers. Species diversity was quantified as species richness using electric-fishing, and environmental conditions were thoroughly described for 81 sites.

We found significant and moderate positive α-SGDCs for all four fish species, whereas β-SGDCs were weaker in strength and positively significant for two of the four species. Causal analyses identified two common variables (geographic isolation and area of available habitats) underlying the α-SGDC relationships. Although weak, we found that β-SGDC correlations related to a direct relationship between taxonomic and genomic differentiation, and to the common influence of the abiotic environment acting as a filter on both species and alleles.

Our study shows that similar ecological and evolutionary processes related to environmental filtering, migration, drift and colonization history act for explaining both species and genetic diversity of fish communities.

II.3 - Introduction

It has long been theoretically acknowledged that parallel processes (i.e. speciation/mutation, ecological/genetic drift, dispersal/gene flow, environmental filtering/natural selection) can shape spatial patterns of species diversity and intraspecific genetic diversity (Antonovics 1976). However, until recently, empirical studies focusing on co-variation in species and intraspecific genetic diversity were rare, which left unanswered the question of whether or not parallel processes can actually shape spatial patterns of species and genetic diversity. To empirically address this issue, Vellend (2003) introduced the Species-Genetic Diversity Correlation concept (SGDC), which quantifies the congruency in the distributions of species and genetic diversity. Since then, the empirical assessment of SGDCs has flourished (reviewed in Vellend et al. 2014). These studies are of critical interest as they (i) test the theoretical statement that similar processes drive patterns of biodiversity at different levels, and (ii) indirectly test for the practical possibility of using one level of diversity as a surrogate for the other for setting conservation plans (He et al. 2008). The general expectation is that species and genetic diversity should co-vary positively, hence producing positive SGDCs (Vellend 2005). Although many SGDC studies confirm this expectation (Vellend et al. 2014), there are still strong discrepancies in the strength and interpretation of SGDCs. Vellend et al. (2014) emphasizes that understanding the bases of these variations should now be a research priority.

SGDC implies a correlative approach, which makes difficult explaining why SGDCs are positive, negative or null, given that "correlation does not imply causation." Several factors can lead to negative or null SGDCs between species and genetic diversity patterns; including opposing evolutionary forces (Derry et al. 2009), factors acting at different temporal or spatial scales (Taberlet et al. 2012), or different responses to static or changing environmental conditions (Puşcaş, Taberlet & Choler 2008). Moreover, under the niche variation hypothesis (Van Valen 1965), an increase in species diversity within a community may reduce the genetic diversity of some species through increased interspecific competition or reduction of the average intraspecific niche breadth (Xu et al. 2016). Contrastingly, three non-exclusive hypotheses have been proposed to explain positive SGDCs (Vellend & Geber 2005). First, species and genetic diversity responding similarly to environmental drivers could show a positive SGDC, such as geographic isolation acting on dispersal and gene flow or habitat area influencing ecological and genetic drift. Second, genetic diversity of one species

may directly increase surrounding community species diversity (Vellend & Geber 2005), by promoting community level stability and reducing extinction risk (Saccheri et al. 1998; Frankham 2015). Third, genetic diversity might increase due to species diversity if increased species diversity generates diversifying selection on non-neutral genetic diversity (Vellend & Geber 2005). Deciphering the relative (or combined) role of each three hypotheses from empirical data is still extremely challenging (Vellend et al. 2014).

All hypotheses described above have been developed by considering correlations between the α component of diversity (Loreau 2000), i.e. between indices of within-sites diversity such as species richness and allelic richness (α-SGDC). However, recent studies also indicate that quantifying between-site diversity (i.e., β-diversity) are important in assessing SGDCs (e.g. Odat, Jetschke & Hellwig 2004; Sei, Lang & Berg 2009; Struebig et al. 2011; Blum et al. 2012). β-diversity measures differences among communities or populations across spatial or temporal scales, thus providing a complementary description of diversity and a more complete understanding of the ecological and evolutionary processes shaping it (Sei et al. 2009; Sexton, Hangartner & Hoffmann 2014).

Vellend (2005) theoretically demonstrated that SGDCs are strongly affected by the abundance of the species from which intraspecific genetic diversity is measured (hereafter, the "target species"), with rarer species producing weaker SGDCs and *vice versa*. In the same vein, Laroche et al. (2015) demonstrated that the mutation-to-gene flow ratio of the target species also strongly affects SGDCs. More specifically, positive SGDCs were theoretically obtained when mutation rate was weak relative to gene flow, whereas SGDCs can be both positive and negative when mutation rate is not negligible (Laroche et al. 2015). Although theoretical works suggest that the strength and sign of SGDCs may vary depending on the target species, most studies assess SGDCs by quantifying genetic diversity from a single species (e.g. Evanno et al. 2009 ; Blum et al. 2012 ; Puscas et al. 2008). However, the peculiarities of each species can provide useful information on the evolutionary and ecological processes shaping diversity, given that life-history traits of species are partaking in shaping spatial patterns of genetic diversity and differentiation (e.g. Duminil et al. 2007; Kelly & Palumbi 2010; Mazé-Guilmo et al. 2016). Studies focusing on comparisons of SGDCs between several species remain scarce (but see Struebig et al. 2011; Taberlet et al. 2012; Lamy et al. 2013; Múrria et al. 2015), although they generally provide key information on the processes underlying SGDCs.

Understanding the processes underlying SGDCs might be particularly tricky in

spatially-structured ecosystems (Blum et al. 2012; Altermatt 2013). This is the case for dendritic river networks (DRNs) in which the dispersal of individuals is highly constrained by the network spatial arrangement (Campbell Grant, Lowe & Fagan 2007; Paz-Vinas & Blanchet 2015). This is especially true for highly water-dependent organisms such as freshwater fish (Paz-Vinas et al. 2015). Moreover, DRNs are highly heterogeneous landscapes, environmentally structured along upstream-downstream gradients (Vannote et al. 1980). Critical environmental components, such as river width and oxygen concentration, vary along stream gradients and are expected to impact the dynamics of species adaptation and selection, thereby having significant impact on SGDCs. These characteristics make the analysis of SGDCs in DRNs challenging, but also highly exciting.

Here, we measured freshwater fish species diversity and intraspecific genetic diversity for four freshwater fish species across the entirety of a river drainage to test (i) if, as expected, the strength and sign of SGDCs vary among the four target species, and (ii) if processes underlying SGDCs are similar among the four target species in a complex DRN. We focused on a set of four contrasting species (*Barbatula barbatula*, *Gobio occitaniae*, *Phoxinus phoxinus*, and *Squalius cephalus*) varying in key biological traits; including rarity and dispersal ability. We predict that SGDCs should be weaker in the less abundant and more vagile species (*S. cephalus*), whereas SGDCs should be stronger for the more abundant and less vagile species (*G. occitaniae*) (see Figure II-1 for specific predictions). Regarding processes underlying SGDCs, we expect processes related to the colonization history of species and populations to have strong and common influences on both genetic and species diversity (Blanchet et al. 2014; Paz-Vinas et al. 2015), and hence on SGDCs. On the contrary, local abiotic parameters such as the level of oxygen concentration or water temperature may have stronger effects on species richness than on genetic diversity, since these parameters have been shown to strongly drive the spatial distribution of freshwater species (Buisson, Blanc & Grenouillet 2008; Blanchet et al. 2014). To test these predictions, we first assessed and compared α-SGDCs and β-SGDCs among the four species at 81 sites covering an entire river drainage. Then, to disentangle the hypotheses proposed by Vellend & Geber (2005), we compiled a detailed environmental and geographical database describing each sampling site, and we applied path analyses (Tenenhaus et al. 2005) for all four target species on both α-SGDCs and β-SGDCs. Path analysis is to our knowledge the most appropriate statistical tool to unravel complex relationships within a set of variables derived from empirical data (Wright 1921). In a companion paper, Seymour et al. (2016) provides complementary findings

of patterns and processes of SGDC in DRNs, but focusing on a taxonomic group (invertebrates) whose dispersal is not restricted to water corridors.

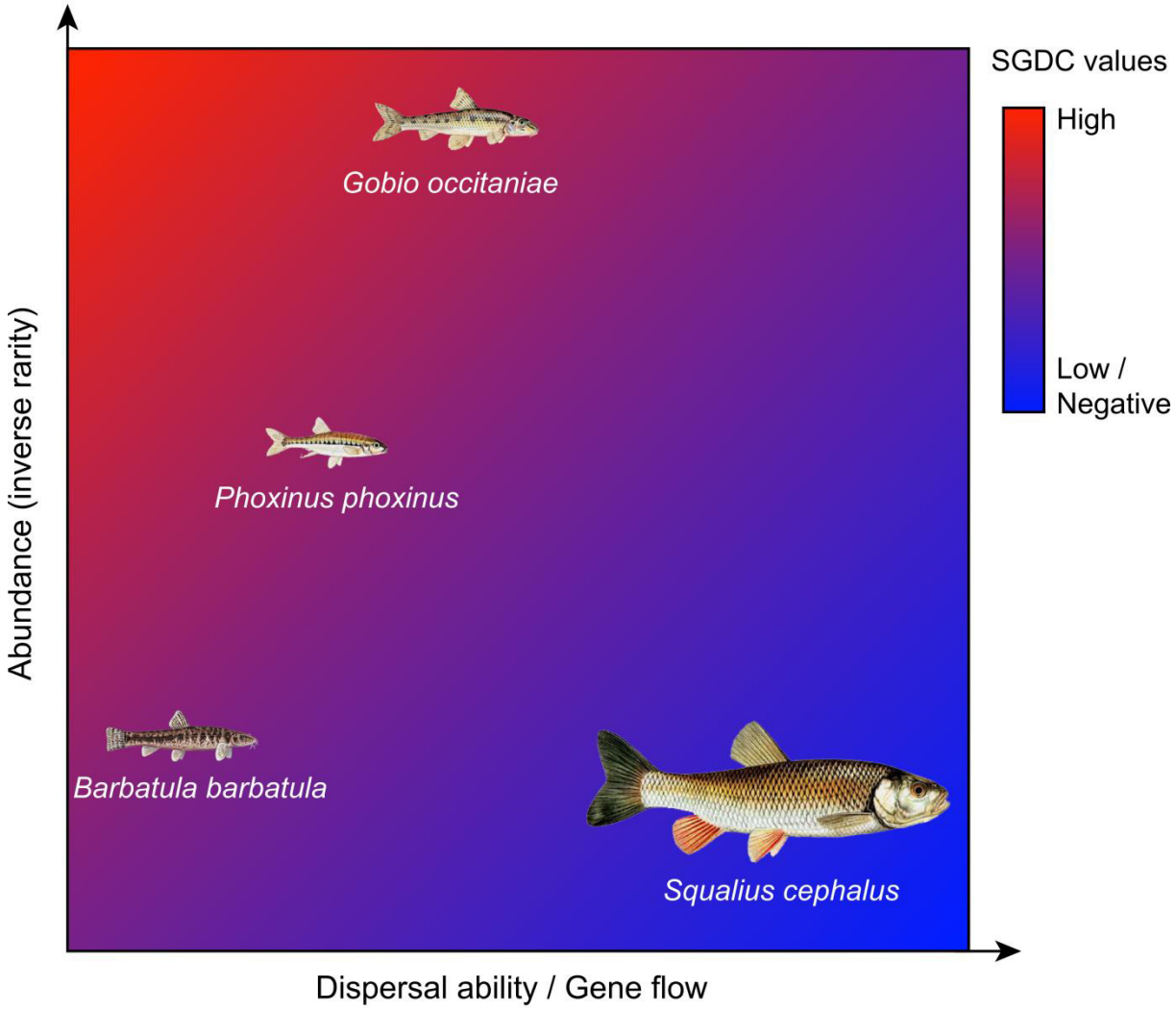


Figure II-1: Theoretical biplot showing how the strength of SGDCs is expected to vary according to the rarity (y-axis) and the gene flow (x-axis) of the target species (i.e. the one used to quantify genetic diversity) and following theoretical works by Vellend (2005) and Laroche *et al.* (2015). We placed on the biplot a picture of each species, which corresponds (based on scientific and our own field knowledge) to their respective level of rarity and dispersal ability (see the main text for details). This allows us to predict that the strength of SGDCs should be different among these four species, and should follow this ranking (from high to low SGDCs): $SGDCs_{G.\ occitaniae} > SGDCs_{P.\ phoxinus} > SGDCs_{B.\ barbatuala} > SGDCs_{S.\ cephalus}$.

II.4 - Materials and methods

Study area

We focused on the Garonne-Dordogne river drainage (South-Western France), that covers an area of 79,800 km². We selected 81 sampling sites evenly scattered across the whole river basin according to the following criteria: (i) sites should be accessible for electric-fishing, (ii) they should host as much of the four target species used for genetic analyses as possible, (iii) taxonomic data should be available, and (iv) sites should cover most of the area of the river basin, and hence most of the environmental variability existing along the upstream-downstream river gradient (see Figure II-2a and Table II-S1).

Species diversity

We collected data on the occurrence of freshwater fish species for all 81 sampling sites using a database provided by the "Office National de l'Eau et des Milieux Aquatiques" (ONEMA; French freshwater agency). The ONEMA yearly monitors fish assemblages in more than 1500 sampling sites in the Garonne-Dordogne river catchment (Poulet, Beaulaton & Dembski 2011), feeding a database gathering presence/absence data for all fish species found in the river basin. These data were collected during electric-fishing campaigns from 1975 to 2011. To describe the taxonomic diversity of each sampling site, we considered data from every yearly sample by site to make the species list as exhaustive as possible. This led to a regional pool of 51 species (see Table II-S2 for details). It is noteworthy that this species list included both native and non-native species (17 non-native species). We ran additional analyses without the non-native species, which led to very similar results (see Table II-S3 for results without the non-native species).

Species α-diversity was quantified as the species richness (i.e. number of species per site), and β-diversity was quantified as the pairwise community dissimilarity using the Jost's index of "true diversity" (Jost 2006), so as to make it directly comparable to genetic β-diversity (see below). This metric measures the community variation among pairs of sites, and ranges from 1 (identical communities) to 2 (completely distinct communities) (Jost 2006). The values were computed using the R package 'simba' (Jurasinski & Retzer 2012).

Genetic diversity

Intraspecific genetic diversity was estimated from four fish species belonging to the

Cypriniformes order: *Gobio occitaniae*, *Phoxinus phoxinus* and *Squalius cephalus* belong to the Cyprinidae family, whereas *Barbatula barbatula* belongs to the Nemacheilidae family. We chose species that are of limited interest for anglers, so as to limit the possibility for past stocking events and uninformed translocations between river drainages. We performed preliminary "outlier population" analyses in our genetic datasets through Factorial Correspondence Analysis using the GENETIX software (Belkhir et al. 1996) and ensured the absence of recent stocking events by asking local angling associations, hence further reducing this risk. Although all these species are mainly insectivorous, they strongly differ in their foraging mode, with *S. cephalus* and *P. phoxinus* feeding in the water column whereas *B. barbatula* and *G. occitaniae* feed preferentially on the bottom. Moreover, these species vary in their level of habitat specialization, with *G. occitaniae* being the most generalist (i.e. it is found almost everywhere in the river basin and in many habitat types) and abundant species, whereas *B. barbatula* is a specialist species living in very specific habitats (mainly riffles in the midstream sections of rivers) at relatively low abundance. *S. cephalus* and *P. phoxinus* are intermediate species; the former is primarily found in downstream sections at relatively low densities, whereas the later is found in upstream sections at relatively high densities (Keith et al. 2011). Additionally, all four fish species strikingly vary in their mean body length, which is often related to dispersal ability in fish (Radinger & Wolter 2014): the largest is *S. cephalus* (300-500mm) followed by *G. occitaniae* (120-150mm), *B. barbatula* (100-120mm) and *P. phoxinus* (80-90mm) (Keith et al. 2011). Due to its body shape *B. barbatula* is considered a poor swimmer, and thus disperser, whereas *S. cephalus* is expected to have the higher dispersal ability because of its large streamline body length. *Gobio occitaniae* and *P. phoxinus* have intermediate dispersal ability. Variability observed for dispersal ability and rarity in these four species covered a non-negligible (although not complete) part of the trait space of the 51 species found in the basin. We summarized in Figure II-1 the rarity (based on abundance and level of habitat specialization) and dispersal ability of each species, which led to theoretical predictions regarding the strength of SGDCs expected for each species.

Specimens were sampled once across each of the 81 sites using electric-fishing between the summer of 2010 and 2011. Average site area was 500 to 1000 m² to ensure the sampling included the full range of local habitat heterogeneity. We sought to capture up to 25 individuals per species per site, although (i) not all species were found in all sites and (ii) not all sites provided 25 individuals due to low site abundances. Sites in which less than 10 individuals were successfully genotyped were removed from the database, leading to 42 sites

for *B. barbatula*, 74 sites for *G. occitaniae*, 54 sites for *P. phoxinus* and 60 sites for *S. cephalus* (Figure II-2), with a total of 5,405 individual genotypes (see Table II-1 for details). For each individual, a small piece of pelvic fin was collected and preserved in 70% ethanol. DNA was extracted using a salt extraction protocol (Aljanabi & Martinez 1997) and individuals were genotyped for eight to ten microsatellite loci [*B. barbatula* (n = 9); *G. occitaniae* (n = 8); *P. phoxinus* (n = 10); *S. cephalus* (n = 10)]. We used 5-20ng of genomic DNA and QIAGEN® Multiplex PCR Kits (Qiagen, Valencia, CA, USA) to perform PCR amplifications. We provide more details on loci, primer concentrations, PCR conditions and multiplex recipes in Table II-S4. The genotyping was conducted on an ABI PRISM™ 3730 Automated Capillary Sequencer (Applied Biosystems, Foster City, CA, USA), and the scoring of allele sizes was done using GENEMAPPER® v.4.0 (Applied Biosystems).

Table II-1: (a) Number of sites sampled, total number of species over all sites and mean and range of species richness and true diversity; (b) number of sites and individuals sampled for all four species and mean and range of allelic richness and Jost's D.

(a) **Species diversity**	Number of sites	Total number of species	Species richness	True diversity
Freshwater fish	81	51	15.100 [4; 31]	1.302 [1; 2]
(b) **Genetic diversity**	Number of sites	Number of individuals	Allelic richness	Jost's D
Barbatula barbatula	42	911	4.427 [2.639; 5.595]	0.383 [-0.010; 0.880]
Gobio occitaniae	74	1770	5.288 [4.047; 6.240]	0.200 [-0.031; 0.627]
Phoxinus phoxinus	54	1324	5.830 [4.258; 6.631]	0.267 [-0.016; 0.781]
Squalius cephalus	60	1400	3.821 [2.477; 5.638]	0.164 [-0.023; 0.526]

We determined the occurrence of null alleles and potential scoring errors with the program MICROCHECKER 2.3 (Van Oosterhout et al. 2004). We tested for departures from Hardy-Weinberg (HW) equilibrium among loci within sites using the R package 'adegenet' v1.4-2 (Jombart 2008). The program GENEPOP v4.0 (Rousset 2008) was used to assess linkage disequilibrium (LD) among loci within sites.

Genetic α-diversity was measured as mean allelic richness using ADZE v1.0 (Szpiech, Jakobsson & Rosenberg 2008). Beta-diversity was quantified as the pairwise genetic differentiation (between-sites) using Jost's D (Jost 2008) using the R package 'mmod' (Winter

2012). Jost's D values range from zero (or a slightly negative value) when there is all alleles present in both populations (no differentiation) to one when there is no alleles in common between populations (strong differentiation). Compared to classical F_{ST} or G_{ST} estimates, Jost's D is preferred as it is independent from the number of alleles sampled in the population, which allows for a better comparison of inter-specific genetic differentiation (Jost 2008).

Finally, spatial distributions of species and genetic α-diversity across the Garonne-Dordogne were described using krigged maps of species and allelic richness using the R package 'fields' (Nychka, Furrer & Sain 2015).

Environmental and geographic data

Environmental variables were collected for each sampling site; including elevation and river width obtained from the French Theoretical Hydrological Network ("Réseau Hydrologique Théorique français", RHT; Pella et al. 2012), along with water temperature (°C), oxygen concentration (mg/L) and saturation (%), suspended matter (mg/L), pH and conductivity (mS/cm) obtained from the database of the Water Information System of the Adour Garonne basin (SIEAG, "Système d'Information sur l'Eau du Bassin Adour Garonne"; http://adour-garonne.eaufrance.fr). We selected environmental variables that were related to key life-history traits of fish species or integrative of many ecosystemic processes. The SIEAG database gathers chemical characteristics of surface water measured several times a year at numerous sites in the river catchment. Only sites where data were available for March, May, July, September and November of 2011 were selected from the SIEAG database. The mean of these five values was calculated to inform the chemical quality of the sites. When sampling location did not overlap perfectly with SIEAG data, the nearest SIEAG site, along the same river and within a 5km radius, was used. Three sampling sites did not have proximal SIEAG sites to gather reliable information, and were hence removed from subsequent analyses. The betweenness centrality value of the first node (i.e. river confluence) upstream of each site was estimated using NetworkX (Hagberg, Schult & Swart 2008). Betweenness centrality is an index of river connectivity that quantifies the positional importance of a node within a network (Freeman 1977; Estrada & Bodin 2008). Topological distance from the outlet (i.e. distance along the river network between a site and the basin outlet) and topological distance between each pair of sites (i.e. distance along the river network between two sites) were computed using QuantumGIS software (QGIS; Quantum GIS Development Team 2015).

Statistical analysis

To quantify and compare the relationships between species and genetic α-diversity indices (i.e. α-SGDCs) we computed Spearman's rank correlations for each species independently. To quantify and compare the relationships between species and genetic β-diversity indices (i.e. β-SGDCs), and because these data are pairwise matrices, we used Mantel tests with 10,000 permutations.

Partial least squares path modeling (PLS-PM; Tenenhaus et al. 2005) was used to unravel the relationships between environmental and geographic variables and α-diversity indices. This method allows simultaneously testing the significance, sign and strength of multiple paths (i.e. relationships, see Tenenhaus et al. 2005) defined within a dataset. For each path, a coefficient is computed to provide the strength and sign of the relationship, as well as a confidence interval using a bootstrap method. We designed a model containing the paths needed to test the hypotheses formulated by Vellend & Geber (2005). As stated above, the first hypothesis stipulates that genetic and species diversity are correlated because of similar effects of environment features. To test this hypothesis, environmental variables were grouped into five latent variables (i.e. variables representing concepts that cannot be measured but are built from several measured variables, see Tenenhaus et al. 2005). Each latent variable corresponded to local characteristics that might influence both levels of α-diversity: Isolation (altitude and distance from the outlet), Connectivity (betweenness), Area (river width), Oxygen (oxygen concentration and saturation) and Water composition (suspended matter, pH, water temperature and conductivity) (see Table II-2 for general expectations and processes linking these variables to genetic and species diversity). We constructed a model in which these five environmental variables were directly linked to both levels of α-diversity (see Figure II-5 for an illustration). As some of these variables are expected to co-vary spatially, we included paths between them when needed (see Table II-S5 for the values of correlation between environmental latent variables). The second and third hypotheses stipulate that one level of diversity influences the other. We tested these two hypotheses by adding a path between allelic richness and species richness. Because of PLS-PM methodological limitations, we could not determine the direction of the path (i.e. from allelic richness to species richness, hypothesis 2; or from species richness to allelic richness, hypothesis 3), it was thus impossible to statistically decipher between these two hypotheses and they were hence grouped into a single hypothesis linking genetic and species diversity, irrespectively of the direction of the link. We used the allelic richness of the four species as a response variable in four

independent models.

Table II-2: General predictions (and underlying processes) regarding the relationships between each latent variable and genetic and species diversity respectively. These predictions arise from general and local knowledge on the influence of each variable on genetic (Paz-Vinas *et al.* 2015) and species (Buisson *et al.* 2008; Blanchet *et al.* 2014) diversity.

Latent variables	Expected influence on genetic diversity	Expected influence on species diversity
Isolation	Isolated populations should have been colonized later and should harbour lower genetic diversity	Isolated communities should have been colonized later and should be taxonomically depauperated
Connectivity	Connectivity should increase gene flow and hence genetic diversity (while reducing genetic differentiation)	Connectivity should increase dispersal among communities, which should led to richer communities with less differentiation
Area	Larger areas could translate into higher effective population size, and hence genetic diversity	Due to the species-area relationship, higher species diversity is expected in sites with larger areas
Oxygen	Oxygen is a strong limiting factors for many fish species, low oxygen concentrations could translate into low effective population sizes and hence low levels of genetic diversity	Oxygen is strongly driving the spatial distribution of fish species
Water composition	Pollution can have strong effects on the effective population size (and hence genetic diversity), most notably in poorly tolerant species such as *P. phoxinus*	Pollution can have strong effect on the persistence of some species, polluted sites have generally lower species richness

We conducted a similar analysis on species and genetic β-diversity data. The design of the models was the same, but species and allelic richness were replaced by Jost's D and true-diversity respectively. Environmental data were converted into pairwise environmental differences between sites, with the exception of the difference in altitude which was computed as the cumulative difference in altitude between sites along the river network (e.g. the total vertical drop to cover from one site to another). The definitions of the latent variables were unchanged with the exception of Isolation, which was defined as the cumulative difference in altitude and the topological distance between sites. The values of correlation between the environmental latent variables are given in Table II-S5. As the PLS-PM approach does not account for the non-independence of pairwise data, we implemented a specific bootstrap procedure, which corrects for the non-independence of pairwise data, as executed in traditional Mantel tests (Legendre & Fortin 2010). Jost's D of the four species was implemented in four independent models to detect similarity and contrast between species. These analyses were conducted using the R package 'plspm' (Sanchez, Trinchera & Russolillo

2015).

Using the values of path coefficients computed from each of these models, we investigated how the magnitudes of the mechanisms driving α-SGDCs and β-SGDCs (i.e. the effect sizes, ES; Nakagawa & Cuthill 2007) differed between species. We then used a meta-analytic approach to test which of the mechanisms stood out as common for all species. Path coefficients values were treated as correlation coefficients from which we computed the Fisher's *Z* effect size (and its associated variance) for each coefficient and each species (see Rosenberg, Adams & Gurevitch 2000; Nakagawa & Cuthill 2007 for formulas). For each path of the models shown in Figure II-5 and Figure II-6 and for each SGDC type independently (α-SGDC and β-SGDC), we then computed the cumulative effect size ($\bar{E}$) and the 95% confidence interval for all pooled species. This procedure allows accounting for the differences in sampling sizes between species when estimating a mean effect size (see Rosenberg et al. 2000), and to test if a path was significant over all species. Finally, we evaluated if the set of effect sizes used to calculate $\bar{E}$ for each path was homogeneous among species or not. The total heterogeneity of a sample (Q_t) was calculated (Rosenberg et al. 2000), and its significance was tested using chi-square statistics.

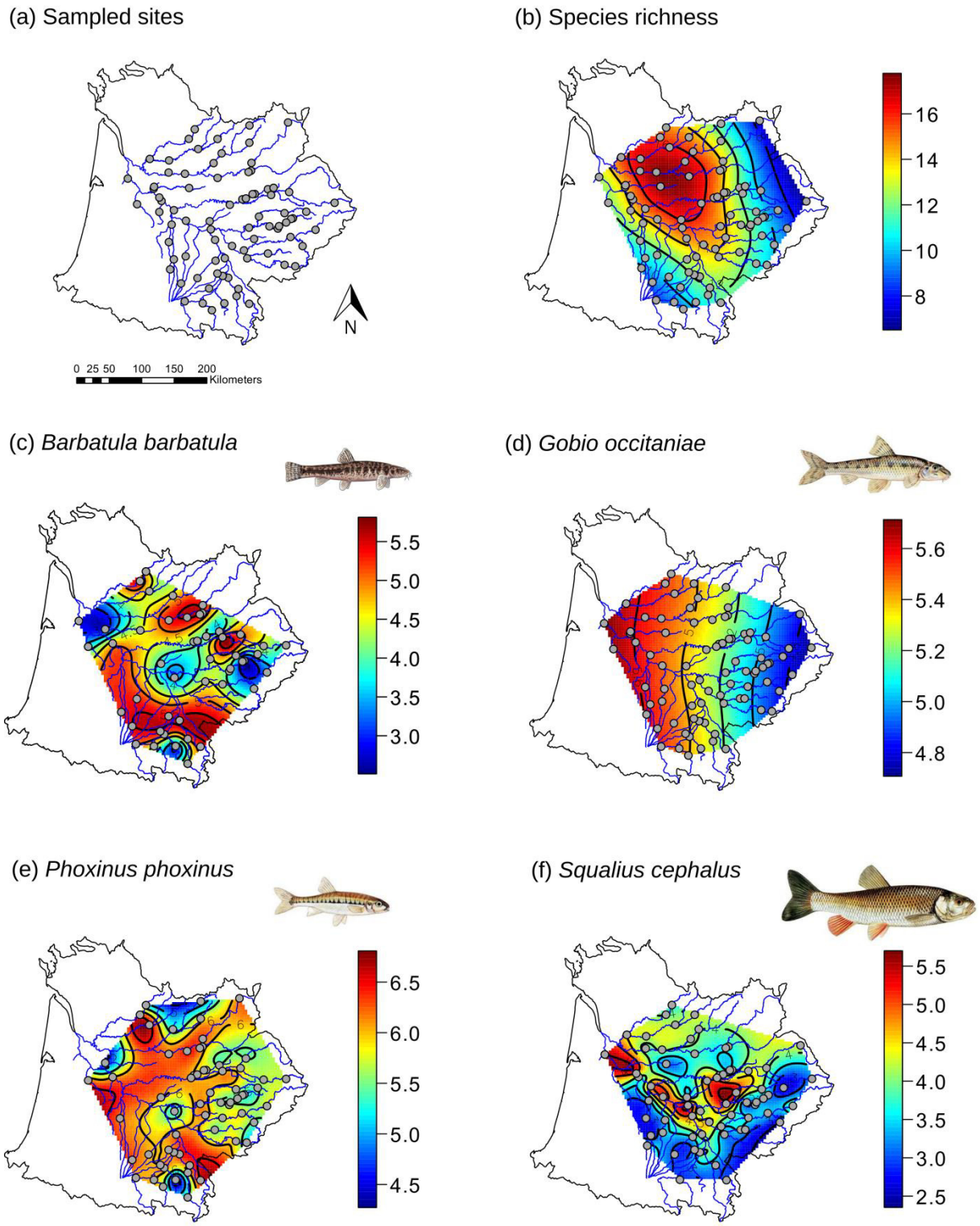


Figure II-2: Maps representing the spatial distribution of interpolated species richness (a) and allelic richness of *B. barbatula* (b), *G. occitaniae* (c), *P. phoxinus* (d) and *S. cephalus* (e). Red color represents high richness and blue color represents low richness. Each grey dot is a sampling site in which the data (species or allelic richness) was available. Interpolated values of β-diversity indices (i.e. true diversity and Jost's D) could not be computed because β-diversity indices take the form of pairwise matrices.

II.5 - Results

Genetic diversity

We found significant deviations from HW, after Bonferroni corrections, for a few locus/population pairs for each species. However, these departures were not consistently clustered among loci or populations for any species (Table II-S6). We also detected significant linkage disequilibrium between a few pairs of loci after Bonferroni corrections for some populations for each species, but no clear patterns appeared for any species across populations (Table II-S6). We did not find evidence for scoring errors due to stuttering or large allele dropouts in our datasets. Given the small extent and random nature of these deviations (HW and linkage disequilibrium), and given the size of the databases, we retained all loci for the subsequent analyses.

Species-genetic diversity correlations

Mean allelic richness ranged from 3.82 ± 0.77 SD (*S. cephalus*) to 5.83 ± 0.48 SD (*P. phoxinus*). The spatial distribution of allelic richness strikingly varied among species (Figure II-2). For instance, in *G. occitaniae*, allelic richness was higher downstream, whereas in *S. cephalus* the highest values of allelic richness were found on main river stretches, near confluence zones. No clear pattern was detected for *B. barbatula* and *P. phoxinus* (Figure II-2). The mean pairwise Jost's D ranged from 0.16 ± 0.09 SD (*S. cephalus*) to 0.38 ± 0.17 SD (*B. barbatula*). Additional genetic α-diversity and β-diversity indices for each species are provided in Table II-1. Mean species richness was 12.77 ± 4.91 SD species and mean true diversity was 1.28 ± 0.12 SD (Table II-1).

We found significant and positive α-SGDCs for each species (Figure II-3). However, correlation coefficients were relatively weak, and contrary to our expectations (Figure II-1), similar among species, ranging from 0.29 to 0.38 (Figure II-3). Similar trends were detected for β-diversity since correlation coefficients between true diversity and Jost's D (β-SGDCs) were weak and of similar strength for all species, although significant and positive β-SGDCs were observed for *B. barbatula* and *G. occitaniae*, but not for *P. phoxinus* and *S. cephalus* (Figure II-4).

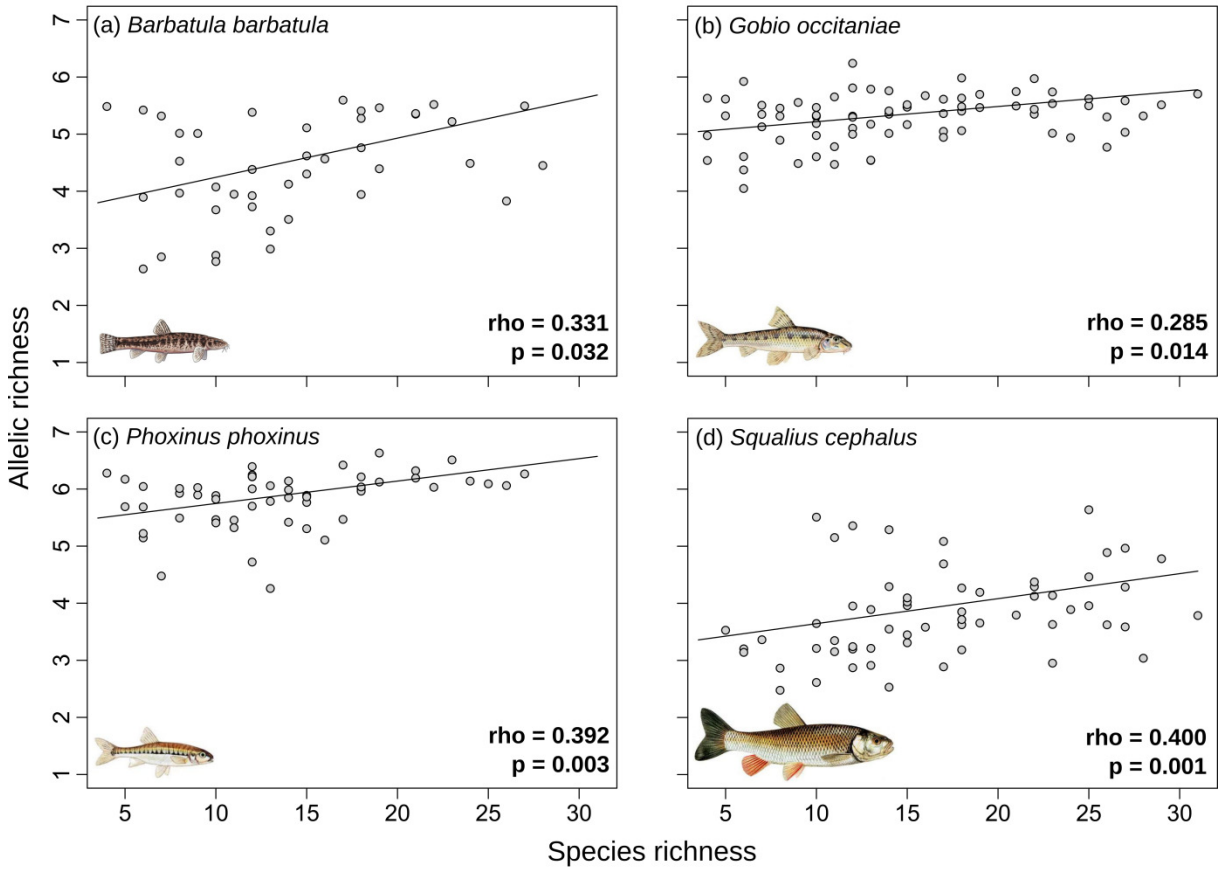


Figure II-3: Allelic richness (genetic α-diversity) of *B. barbatula* (a), *G. occitaniae* (b), *P. phoxinus* (c) and *S. cephalus* (d) plotted against species richness (species α-diversity) with Spearman's rho and associated p-values.

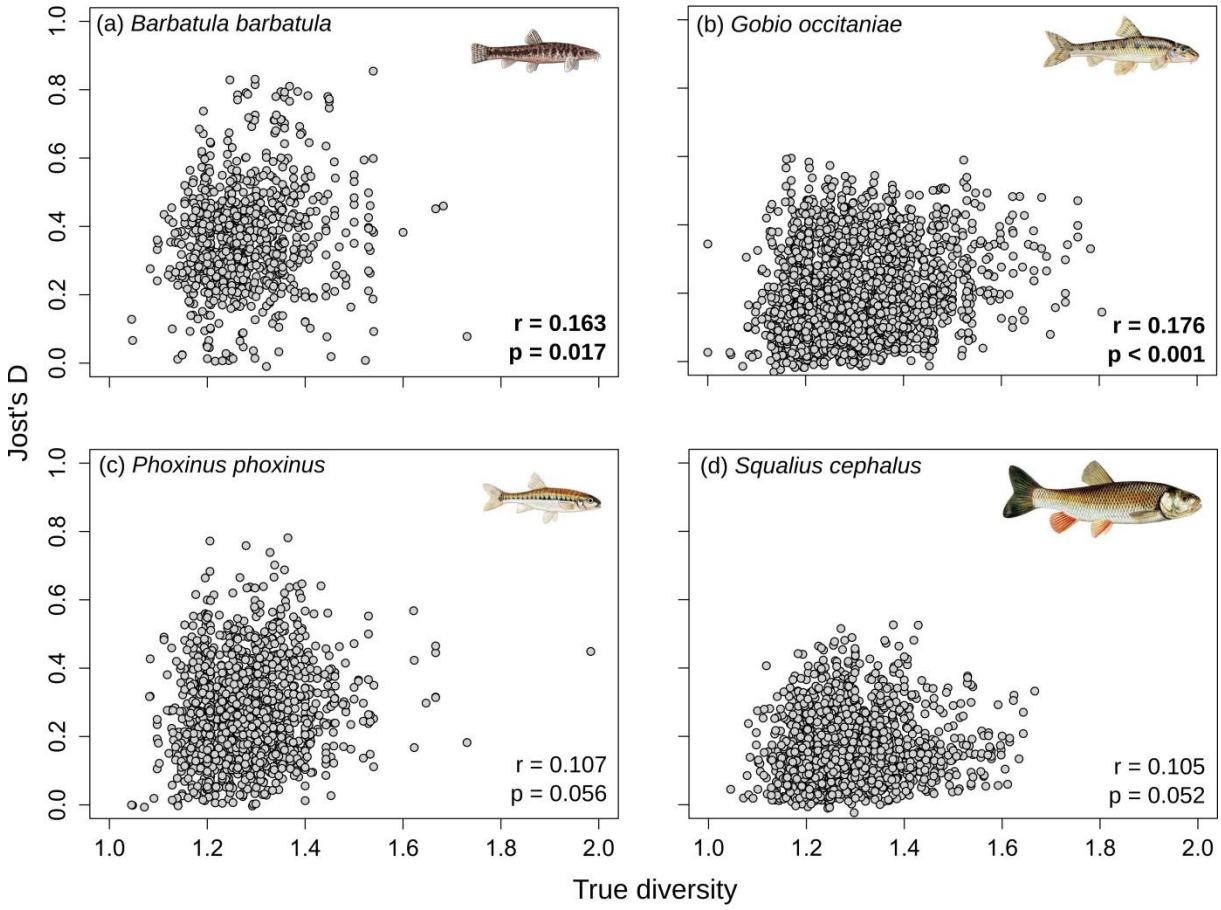


Figure II-4: Jost's D (genetic β-diversity) of *B. barbatula* (a), *G. occitaniae* (b), *P. phoxinus* (c) and *S. cephalus* (d) plotted against species richness (species β-diversity) with Mantel's r and associated p-values.

Meta-analytic approach on partial least squares path modeling

For all species two variables significantly affected species and genetic α-diversity. First, the isolation of a site (i.e. altitude and distance from the outlet) was negatively related to both levels of α-diversity (Figure II-5; Table II-3). The magnitudes of isolation effects on species α-diversity and genetic α-diversity were significantly different from zero (Table II-3). The effect of isolation on genetic α-diversity was significantly heterogeneous among species (Q_t = 19.86, d.f. = 3, P = 0.001), mainly because the effect of isolation on genetic α-diversity was stronger in *G. occitaniae* (ES = -0.91) than in any other species (ESs ≤ -0.29, see Table II-S7 for details). Second, the cumulative effect of available area on species α-diversity and genetic α-diversity species was significantly positive for all species (Table II-3) and homogeneous among species (Q_t = 2.93, d.f. = 3, P = 0.40 and Q_t = 6.67, d.f. = 3, P = 0.08 respectively). We found neither evidence for additional environmental effects on α-diversity,

nor significant relationships between species richness and allelic richness (Table II-3, Figure II-5).

Considering β-diversity, we found significant effect sizes between Jost's D and true diversity for all species (Table II-3). However, we were not able to determine the direction of the arrow (i.e. from Jost's D to true diversity or from true diversity to Jost's D) because of PLS-PM methodological limitations. This effect was heterogeneous among species (Q_t = 10.25, d.f. = 3, P = 0.02), with higher values for *B. barbatula* and *G. occitaniae* (ESs = 0.11 and 0.12 respectively) than for *P. phoxinus* and *S. cephalus* (ESs = 0.02 and 0.05 respectively). Further, the difference in water composition between sites was significantly and positively related to both levels of β-diversity for all species (Table II-3, Figure II-6). The isolation and area of sites were drivers of genetic β-diversity (Table II-3), although their effects were heterogeneous among species (both P < 0.001). Finally, difference in water oxygenation had a negative effect on species β-diversity (Table II-3), however this effect was heterogeneous among species (Q_t = 15.15, d.f. = 3, P = 0.002).

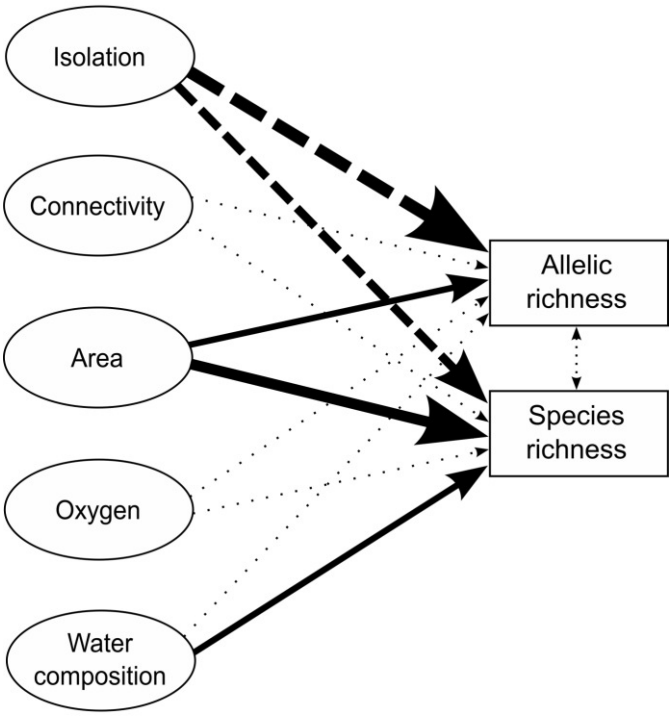


Figure II-5: Graphical representation of the meta-analysis results obtained from Partial least square path modeling on α-diversity indices over all species. Hatched arrows correspond to significant negative mean effect sizes and plain arrows correspond to significant positive mean effect sizes; their size is proportional to the absolute value of their mean effect sizes. Thin dotted arrows represent non-significant mean effect sizes. Variable names enclosed in squares depict observed variables while variable names enclosed in circles depict latent variables. For the sake of clarity, the paths linking environmental variables are not shown.

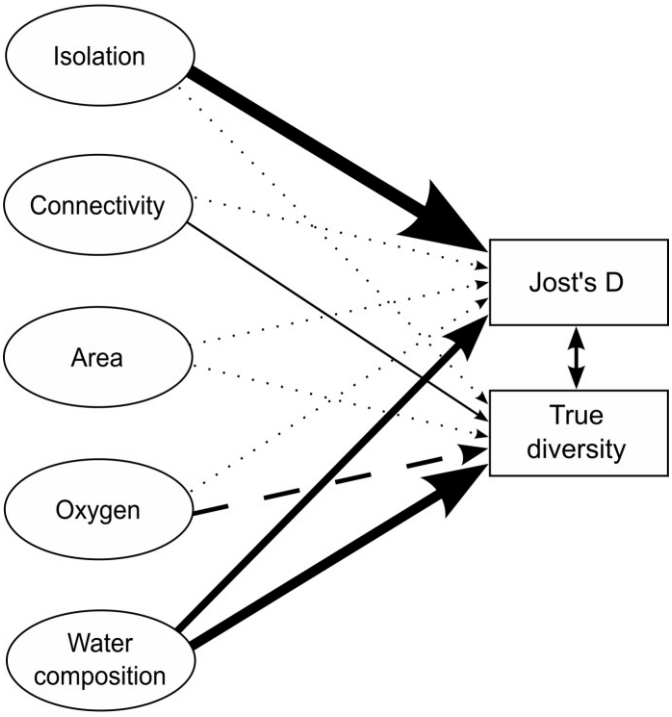


Figure II-6: Graphical representation of the meta-analysis results obtained from Partial least square path modeling on β-diversity indices over all species. See Legend in Figure II-5 for details.

Table II-3: Mean effect size ($\bar{E}$), 95% confidence interval (95% CI), total heterogeneity of a sample (Q_t) and corresponding p-value (P) computed from the path coefficients obtained from partial least-square path modeling applied to (a) α-diversity indices and (b) β-diversity indices. Values in bold are significant mean effect sizes and the corresponding 95% confidence intervals.

(a) **α-diversity**	$\bar{E}$	95% CI	Q_t	P
Isolation → Allelic richness	**-0.464**	**[-0.684; -0.244]**	19.863	< 0.001
Area → Allelic richness	**0.238**	**[0.017; 0.458]**	6.647	0.084
Connectivity → Allelic richness	0.079	[-0.141; 0.299]	1.767	0.622
Water composition → Allelic richness	0.121	[-0.099; 0.341]	3.245	0.355
Oxygen → Allelic richness	0.175	[-0.045; 0.395]	3.628	0.304
Isolation → Species richness	**-0.322**	**[-0.542; -0.102]**	0.352	0.950
Area → Species richness	**0.405**	**[0.185; 0.625]**	3.354	0.340
Connectivity → Species richness	-0.089	[-0.309; 0.131]	0.850	0.837
Water composition → Species richness	**0.238**	**[0.018; 0.458]**	3.822	0.281
Oxygen → Species richness	0.175	[-0.395; 0.045]	0.382	0.944
Allelic richness → Species richness	-0.012	[-0.232; 0.208]	0.755	0.862
(b) **β-diversity**	$\bar{E}$	95% CI	Q_t	P
Isolation → Jost's D	**0.316**	**[0.276; 0.356]**	344.950	< 0.001
Area → Jost's D	-0.025	[-0.065; 0.015]	15.381	0.001
Connectivity → Jost's D	0.019	[-0.021; 0.059]	63.807	< 0.001
Water composition → Jost's D	**0.175**	**[0.136; 0.215]**	240.067	< 0.001
Oxygen → Jost's D	-0.018	[-0.058; 0.022]	38.508	< 0.001
Isolation → True diversity	-0.011	[-0.051; 0.029]	34.910	< 0.001
Area → True diversity	-0.035	[-0.075; 0.004]	6.416	0.093
Connectivity → True diversity	**0.053**	**[0.013; 0.092]**	6.319	0.097
Water composition → True diversity	**0.235**	**[0.195; 0.275]**	47.929	< 0.001
Oxygen → True diversity	**-0.107**	**[-0.147; -0.067]**	5.621	0.132
Jost's D → True diversity	**0.086**	**[0.046; 0.126]**	12.639	0.005

II. 6 - Discussion

α-SGDCs in riverine fish species

Contrary to our theoretical expectations (Figure II-1), α-SGDCs were significantly positive and similar in statistical strength for all four target species. Most previous empirical

studies involving multi-species approaches have highlighted the importance of high inter-specific variability in sign and strength of α-SGDCs among species (Struebig et al. 2011; Taberlet et al. 2012 but see Lamy et al. 2013). Additionally, high inter-specific variability in α-SGDCs has recently been theoretically explained (Laroche et al. 2015). However, our results are consistent with previous findings showing that similar positive α-SGDCs are detected when species and genetic α-diversity are extracted from ecologically similar taxonomic groups (He & Lamont 2010; Seymour et al. 2016). For instance, in a companion paper involving invertebrates across a large river basin, Seymour et al. (2016) found that α-SGDC was stronger when intraspecific genetic α-diversity of *Gammarus sp*. was compared to species α-diversity of Amphipoda than to other invertebrate taxa such as Ephemeroptera, Plecoptera or Trichoptera. This suggests that the four target species we used to estimate genetic α-diversity may not be different enough to generate α-SGDCs of different strength or sign. This is rather surprising given that these species occupy very different ecological niches and strongly vary in abundance and for many biological traits (Keith et al. 2011), with some of these traits having previously been shown to affect the strength and sign of SGDCs (Vellend 2005; Laroche et al. 2015). Nonetheless, the use of path analysis in our study, combined with a meta-analytic approach, demonstrated that such similar α-SGDCs arose from common mechanisms involving geographic isolation and the area of the sampled habitat.

We demonstrated that hypotheses linking species and genetic richness through a causal relationship were discarded for all species. On the contrary, we found that similar effects of environmental variables on both levels of α-diversity (hypothesis 1 of Vellend & Geber 2005) were the main drivers of positive α-SGDCs. First, the geographic isolation of sites negatively impacted both levels of α-diversity, such that species and allelic richness tended to be lower in high altitude sites, located far away from the outlet. These negative effects of isolation on α-diversity are consistent with previous findings (Buisson et al. 2008; Paz-Vinas et al. 2015), and may be explained with two non-mutually exclusive mechanisms. First, sites at low altitude and geographically close to river outlets are expected to experience greater immigration (i.e. sink populations; Gotelli & Taylor 1999; Cadotte 2006; Paz-Vinas et al. 2015), which might provide both new species and novel alleles to these sites, thus counteracting the effect of drift and increasing α-diversity (Vellend & Geber 2005). Second, high altitude and increased distance from river outlets can reflect colonization gradients, resulting from historical glacial refugia (Costedoat & Gilles 2009; Blanchet et al. 2014; Paz-Vinas et al. 2015), which likely contributes to species and genetic α-diversity gradients. For

instance, Paz-Vinas et al. (2015) demonstrated that the downstream increase in genetic α-diversity generally observed in rivers was mainly due to past colonization pathways from downstream to upstream in many freshwater organisms. Additionally, we found that site area (i.e. the river width) positively impacted both levels of α-diversity. Sites of higher river width are expected to sustain high fish density, thus increasing (i) the effective population size of a species, which is known to correlate positively with genetic α-diversity (Frankham 1996; Raeymaekers et al. 2008) and (ii) the number of species that can cohabit a site (Jackson, Peres-Neto & Olden 2001). Furthermore, large stretches of rivers are expected to display greater environmental heterogeneity, which might increase species and genetic α-diversity through diversifying selection (Vellend & Geber 2005).

β-SGDCs in riverine fish species

β-SGDCs were weaker in strength than α-SGDCs, although positive and within the same order of magnitude among species. This result was unexpected since most SGDCs studies investigating both α- and β-diversity found -in average- stronger β-SGDCs than α-SGDCs (see for example Odat et al. 2004; Sei et al. 2009).

We provided strong evidence that mechanisms driving β-SGDCs differ to some extent from those driving α-SGDCs. We found evidence that hypotheses linking genetic differentiation to species differentiation could not be discarded. Although, we were not able to tease apart the direction (genetic to taxonomy, or taxonomy to genetic) of the relationship between species and genetic β-diversity, two non-exclusive mechanisms can be considered. Species β-diversity may influence genetic β-diversity if taxonomically distinct communities generate differential selective pressures acting on local populations, hence favoring different genotypes (Vellend & Geber 2005). Given that we focused on neutral genetic markers, therein observing evidence of neutral population divergence, this first explanation is unlikely to explain our results. It is now well acknowledged that genetically-differentiated populations, even at small spatial scales, can influence differential structure among communities and ecosystems (e.g. Harmon et al. 2009; Bassar et al. 2010). Therefore, a second explanation is that genetic drift led to morphological, physiological or behavioural divergences among populations thereby influencing the structure of the local ecosystems and communities. Either way, our work suggests that population and community differentiation are interrelated. An important next step will be to further unravel the exact mechanisms underlying the relationships between different levels of diversity.

We further showed that hypothesis 1 from Vellend & Geber (2005) was also likely to explain observed patterns, since differences in water composition had a positive influence on both genetic and species β-diversity. This result might reflect a mechanism of isolation-by-environment (Wang & Bradburd 2014; Sexton et al. 2014), by which gene flow between environmentally different sites is limited by the maladaptation of immigrants. Although this mechanism strongly relies on the effects of local selective pressures, it can also affect neutral genetic diversity when gene flow is reduced between sites differing in their local environments (Sexton et al. 2014). Similarly, the structure and composition of a community is expected to vary with environmental features, here water composition, when species show strong habitat preferences. At the community level, environmental filtering may strongly drive community composition, which has been demonstrated in diverse freshwater organisms, including fish (Blanchet et al. 2014; Heino, Melo & Bini 2015). Environmental filtering might thus occur both at the population and community levels to drive co-variation between community and population compositions.

In addition to observing a common influence of environmental filtering on species and genetic diversity, we also found that genetic and species β-diversity were driven by independent processes. For instance, our results suggest that genetic β-diversity was also driven by isolation-by-distance for all four species, which was not the case for species β-diversity. This shows that the equilibrium between genetic drift and gene flow in these four species is affected by geographical isolation, although not similarly as the effect size of isolation on genetic differentiation was heterogeneous among species. Because species β-diversity was not impacted by isolation, community and population differentiations may be partly driven by independent evolutionary and ecological processes, which may explain why β-SGDCs tended to be weaker than α-SGDCs. Similarly, the heterogeneity in the effects of geographic isolation on genetic differentiation among species may explain why β-SGDCs vary among species (in terms of strength and significance).

The role of spatial extent in quantifying SGDCs

It is noteworthy that we observed heterogeneous effect sizes among the four species for several pathways explaining species α- and β-diversity. This is a rather unexpected outcome as the species lists were extracted from a unique database, and we expected species α- and β-diversity indices to be related to the same variables, irrespectively of the target species. This heterogeneity could result from spatial-scale effects. Indeed, although the target

populations showed a relatively high degree of spatial congruence, there were noticeable differences in spatial extents and environmental variabilities for the datasets used for each target species. These differences in spatial extents may explain the heterogeneity in effect sizes of several environmental variables on species α- and β-diversity between target species (Lira-Noriega et al. 2007; Cushman & Landguth 2010). However, these spatial extent effects seemed to be of minor consequences and did not prevent the identification of the main mechanisms driving SGDCs as significant mean effect sizes were identified.

Conclusion

We demonstrated that combined effects of isolation and habitat area at the population and community levels are likely to be responsible for the positive α-SGDCs observed in the four freshwater fish species. We further showed that positive β-SGDCs are likely to be generated by two co-occurring mechanisms, although additional mechanisms uniquely affecting one of the two levels of β-diversity probably contribute to make β-SGDCs weaker than α-SGDCs. To sum up, while species and genetic α-diversity were likely to be driven by similar landscape features, it appeared that species and genetic β-diversity were underlined by a combination of common and unique eco-evolutionary processes.

Alpha and β-SGDCs often vary strongly among studied organisms, even in a single landscape (e.g. Struebig et al. 2011; Taberlet et al. 2012), which was unexpectedly not verified in our study. We believe that the next important challenge of SGDC studies will be to better quantify the level of variation in life-history traits needed to generate different (or similar) α- and β-SGDCs among species. To solve this issue empirically it seems relevant to focus on several target species, although researchers will need to carefully design the study to encompass large ranges of life-history strategies (as in Taberlet et al. 2012). The meta-analytic approach we used allowed highlighting both common and unique mechanisms driving α- and β-diversity in these four species. All four fish species we considered here are from a similar trophic level (secondary consumers), but they varied strikingly in many traits. From our results, we can conclude that dissimilarities in species traits have no or little impact on the two mechanisms driving α-SGDCs, namely the effects of isolation and habitat area, whereas they may generate slight dissimilarities in the mechanisms driving genetic β-diversity.

II.7 - References

Aljanabi S.M. & Martinez I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research* **25**, 4692–4693.

Altermatt F. (2013) Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology* **47**, 365–377.

Antonovics J. (1976) The Input from Population Genetics: "The New Ecological Genetics." *Systematic Botany* **1**, 233–245.

Bassar R.D., Marshall M.C., López-Sepulcre A., Zandonà E., Auer S.K., Travis J., *et al.* (2010) Local adaptation in Trinidadian guppies alters ecosystem processes. *Proceedings of the National Academy of Sciences* **107**, 3616–3621.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Blanchet S., Helmus M.R., Brosse S. & Grenouillet G. (2014) Regional vs local drivers of phylogenetic and species diversity in stream fish communities. *Freshwater Biology* **59**, 450–462.

Blum M.J., Bagley M.J., Walters D.M., Jackson S.A., Daniel F.B., Chaloud D.J., *et al.* (2012) Genetic diversity and species diversity of stream fishes covary across a land-use gradient. *Oecologia* **168**, 83–95.

Buisson L., Blanc L. & Grenouillet G. (2008) Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish* **17**, 244–257.

Cadotte M.W. (2006) Dispersal and Species Diversity: A Meta-Analysis. *The American Naturalist* **167**, 913–924.

Campbell Grant E.H., Lowe W.H. & Fagan W.F. (2007) Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters* **10**, 165–175.

Costedoat C. & Gilles A. (2009) Quaternary Pattern of Freshwater Fishes in Europe: Comparative Phylogeography and Conservation Perspective. *The Open Conservation Biology Journal* **3**, 36–48.

Cushman S.A. & Landguth E.L. (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology* **19**, 3592–3602.

Derry A.M., Arnott S.E., Shead J.A., Hebert P.D.N. & Boag P.T. (2009) Ecological linkages

between community and genetic diversity in zooplankton among boreal shield lakes. *Ecology* **90**, 2275–2286.

Duminil J., Fineschi S., Hampe A., Jordano P., Salvini D., Vendramin G.G., *et al.* (2007) Can Population Genetic Structure Be Predicted from Life-History Traits? *The American Naturalist* **169**, 662–672.

Estrada E. & Bodin Ö. (2008) Using network centrality measures to manage landscape connectivity. *Ecological Applications* **18**, 1810–1825.

Frankham R. (2015) Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* **24**, 2610–2618.

Frankham R. (1996) Relationship of Genetic Variation to Population Size in Wildlife. *Conservation Biology* **10**, 1500–1508.

Freeman L. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry* **40**, 35–41.

Gotelli N.J. & Taylor C.M. (1999) Testing metapopulation models with stream-fish assemblages. *Evolutionary Ecology Research* **1**, 835–845.

Hagberg A.A., Schult D.A. & Swart P.J. (2008) Exploring Network Structure, Dynamics, and Function using NetworkX. In: *Proceedings of the 7th Python in Science Conference (ScyPy2008)*. (Eds G. Varoquaux, T. Vaught & J. Millman), pp. 11–15. Pasadena, CA.

Harmon L.J., Matthews B., Des Roches S., Chase J.M., Shurin J.B. & Schluter D. (2009) Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* **458**, 1167–1170.

He T. & Lamont B.B. (2010) Species versus genotypic diversity of a nitrogen-fixing plant functional group in a metacommunity. *Population Ecology* **52**, 337–345.

He T., Lamont B.B., Krauss S.L., Enright N.J. & Miller B.P. (2008) Covariation between intraspecific genetic diversity and species diversity within a plant functional group: Covariation between species and genetic diversity. *Journal of Ecology* **96**, 956–961.

Heino J., Melo A.S. & Bini L.M. (2015) Reconceptualising the beta diversity-environmental heterogeneity relationship in running water systems. *Freshwater Biology* **60**, 223–235.

Jackson D.A., Peres-Neto P.R. & Olden J.D. (2001) What controls who is where in freshwater fish communities - the roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 157–170.

Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405.

Jost L. (2006) Entropy and diversity. *Oikos* **113**, 363–375.

Jost L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015–4026.

Jurasinski G. & Retzer V. (2012) *simba: A Collection of functions for similarity analysis of vegetation data*.

Keith P., Persat H., Feunteun E., Adam B. & Geniez M. (2011) *Les poissons d'eau douce de France*. Muséum National d'Histoire Naturelle; Biotope, Paris; Mèze.

Kelly R.P. & Palumbi S.R. (2010) Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLOS ONE* **5**, e8594.

Lamy T., Jarne P., Laroche F., Pointier J.-P., Huth G., Segard A., *et al.* (2013) Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. *Molecular Ecology* **22**, 4445–4456.

Laroche F., Jarne P., Lamy T., David P. & Massol F. (2015) A Neutral Theory for Interpreting Correlations between Species and Genetic Diversity in Communities. *The American Naturalist* **185**, 59–59.

Legendre P. & Fortin M.-J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* **10**, 831–844.

Lira-Noriega A., Soberón J., Navarro-Sigüenza A.G., Nakazawa Y. & Peterson A.T. (2007) Scale dependency of diversity components estimated from primary biodiversity data and distribution maps. *Diversity and Distributions* **13**, 185–195.

Loreau M. (2000) Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* **91**, 3–17.

Mazé-Guilmo E., Blanchet S., McCoy K.D. & Loot G. (2016) Host dispersal as the driver of parasite genetic structure: a paradigm lost? *Ecology Letters* **19**, 336–347.

Múrria C., Rugenski A.T., Whiles M.R. & Vogler A.P. (2015) Long-term isolation and endemicity of Neotropical aquatic insects limit the community responses to recent amphibian decline. *Diversity and Distributions* **21**, 938–949.

Nakagawa S. & Cuthill I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **82**, 591–605.

Nychka D., Furrer R. & Sain S. (2015) *fields: Tools for Spatial Data*.

Odat N., Jetschke G. & Hellwig F.H. (2004) Genetic diversity of *Ranunculus acris* L. (Ranunculaceae) populations in relation to species diversity and habitat type in grassland

communities. *Molecular Ecology* **13**, 1251–1257.

Paz-Vinas I. & Blanchet S. (2015) Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology* **28**, 986–994.

Paz-Vinas I., Loot G., Stevens V.M. & Blanchet S. (2015) Evolutionary processes driving spatial patterns of intra-specific genetic diversity in river ecosystems. *Molecular Ecology* **24**, 4586–4604.

Pella H., Lejot J., Lamouroux N. & Snelder T. (2012) Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. *Géomorphologie : relief, processus, environnement*, 317–336.

Poulet N., Beaulaton L. & Dembski S. (2011) Time trends in fish populations in metropolitan France: insights from national monitoring data. *Journal of Fish Biology* **79**, 1436–1452.

Puşcaş M., Taberlet P. & Choler P. (2008) No positive correlation between species and genetic diversity in European alpine grasslands dominated by *Carex curvula*. *Diversity and Distributions* **14**, 852–861.

Quantum GIS Development Team (2015) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project.

Radinger J. & Wolter C. (2014) Patterns and predictors of fish dispersal in rivers. *Fish and Fisheries* **15**, 456–473.

Raeymaekers J.A.M., Maes G.E., Geldof S., Hontis I., Nackaerts K. & Volckaert F.A.M. (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. *Evolutionary Applications* **1**, 475–488.

Rosenberg M.S., Adams D.C. & Gurevitch J. (2000) *MetaWin: statistical software for meta-analysis*, Version 2.0. Sinauer Associates, Sunderland, MA.

Rousset F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106.

Saccheri I., Kuussaari M., Kankare M., Vikman P., Fortelius W. & Hanski I. (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**, 491–494.

Sanchez G., Trinchera L. & Russolillo G. (2015) *plspm: Tools for Partial Least Squares Path Modeling (PLS-PM)*.

Sei M., Lang B.K. & Berg D.J. (2009) Genetic and community similarities are correlated in endemic-rich springs of the northern Chihuahuan Desert. *Global Ecology and Biogeography* **18**, 192–201.

Sexton J.P., Hangartner S.B. & Hoffmann A.A. (2014) Genetic Isolation by Environment or

Distance: Which Pattern of Gene Flow Is Most Common? *Evolution* **68**, 1–15.

Seymour M., Seppälä K., Mächler E. & Altermatt F. (in review) Lessons from the macroinvertebrates: species-genetic diversity correlations highlight important dissimilar relationships. *Freshwater biology*

Struebig M.J., Kingston T., Petit E.J., Le Comber S.C., Zubaid A., Mohd-Adnan A., *et al.* (2011) Parallel declines in species and genetic diversity in tropical forest fragments: Parallel diversity declines in forest fragments. *Ecology Letters* **14**, 582–590.

Szpiech Z.A., Jakobsson M. & Rosenberg N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504.

Taberlet P., Zimmermann N.E., Englisch T., Tribsch A., Holderegger R., Alvarez N., *et al.* (2012) Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters* **15**, 1439–1448.

Tenenhaus M., Vinzi V.E., Chatelin Y.-M. & Lauro C. (2005) PLS path modeling. *Computational Statistics & Data Analysis* **48**, 159–205.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538.

Van Valen L. (1965) Morphological Variation and Width of Ecological Niche. *The American Naturalist* **99**, 377–390.

Vannote R.L., Minshall G.W., Cummins K.W., Sedell J.R. & Cushing C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences* **37**, 130–137.

Vellend M. (2003) Island Biogeography of Genes and Species. *The American Naturalist* **162**, 358–365.

Vellend M. (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *The American Naturalist* **166**, 199–215.

Vellend M. & Geber M.A. (2005) Connections between species diversity and genetic diversity. *Ecology Letters* **8**, 767–781.

Vellend M., Lajoie G., Bourret A., Múrria C., Kembel S.W. & Garant D. (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology* **23**, 2890–2901.

Wang I.J. & Bradburd G.S. (2014) Isolation by environment. *Molecular Ecology* **23**, 5649–5662.

Wright S. (1921) Correlation and Causation. *Journal of agricultural research* **20**, 557–585.

Xu W., Liu L., He T., Cao M., Sha L., Hu Y., *et al.* (2016) Soil properties drive a negative correlation between species diversity and genetic diversity in a tropical seasonal rainforest. *Scientific Reports* **6**.

II.8 - Acknowledgments

We are very grateful to all the ONEMA staff that contributed to the field sampling. Leslie Faggiano, Charlotte Evangelista, Christine Lauzeral, Loïc Tudesque, Roselyne Etienne, Sébastien Villéger, Julien Cucherousset, Vincent Dubut, Charlotte Veyssière and Alain Blanchet. Charlotte Veyssière, Roselyne Etienne and Vincent Dubut are also thanked for their help with molecular analyses. We warmly thank Mathew Seymour and Florian Altermatt for constructive discussions and for commenting/correcting the manuscript. Two reviewers provided insightful comments that greatly improve the quality of the manuscript. LF was financially supported by a MESR ("Ministère de l'Enseignement Supérieur et de la Recherche") PhD scholarship during this study. This project has been carried out with financial support from the Commission of the European Communities, specific RTD programme "IWRMNET". This work has been done in two research units (EDB & SETE) that are part of the "Laboratoire d'Excellence" (LABEX) entitled TULIP (ANR-10-LABX-41).

II.9 - Supplementary materials for Chapter II

Table II-S1: Minimum, mean, maximum and standard deviation of environmental variables.

Environmental variables	Minimum	Mean	Maximum	Standard deviation
Elevation	6.90	211.65	751.50	171.97
Width	4.29	37.84	46.11	31.97
Temperature	8.92	14.10	17.86	2.10
Oxygen concentration	7.20	9.78	12.03	0.81
Oxygen saturation	72.12	96.35	118.12	7.05
Suspended matter	2.00	11.15	90.40	12.91
pH	7.07	7.91	8.66	0.33
Conductivity	51.00	247.73	581.80	128.15

Table II-S2: List of the 51 species of the regional pool, their family and percentage of occurrence over the 81 sites. Species in bold are the four target species of the study.

Family	Species	Occurrence
Anguillidae	*Anguilla anguilla*	60.49%
Blennidae	*Salaria fluviatilis*	1.23%
Centrarchidae	*Lepomis gibbosus*	48.14%
Centrarchidae	*Micropterus salmoides*	19.75%
Clupeidae	*Alosa alosa*	2.46%
Cottidae	*Cottus gobio*	23.45%
Cyprinidae	*Abramis brama*	29.62%
Cyprinidae	*Alburnoïdes bipunctatus*	2.46%
Cyprinidae	*Alburnus alburnus*	72.83%
Cyprinidae	*Barbus barbus*	83.95%
Cyprinidae	*Barbus meridionalis*	2.46%
Cyprinidae	*Blicca bjoerkna*	32.9%
Cyprinidae	*Carassius auratus*	4.93%
Cyprinidae	*Carassius carassius*	35.80%
Cyprinidae	*Ctenopharyngodon idella*	1.23%
Cyprinidae	*Cyprinus carpio*	43.20%
Cyprinidae	***Gobio occitaniae***	97.53%
Cyprinidae	*Hypophthalmichthys molitrix*	8.64%
Cyprinidae	*Leucaspius delineatus*	1.23%
Cyprinidae	*Leuciscus leuciscus*	76.54%
Cyprinidae	*Pachychilon pictum*	16.4%
Cyprinidae	*Parachondrostoma toxostoma*	45.67%
Cyprinidae	***Phoxinus phoxinus***	83.95%
Cyprinidae	*Pseudorasbora parva*	13.58%
Cyprinidae	*Rhodeus amarus*	18.51%
Cyprinidae	*Rutilus rutilus*	77.77%
Cyprinidae	*Scardinius erythrophtalmus*	30.86%
Cyprinidae	***Squalius cephalus***	92.59%
Cyprinidae	*Telestes souffia*	1.23%
Cyprinidae	*Tinca tinca*	40.74%
Esocidae	*Esox lucius*	40.74%
Gasterosteidae	*Gasterosteus gymnurus*	3.70%
Gasterosteidae	*Pungitius laevis*	2.46%
Ictaluridae	*Ameiurus melas*	20.98%
Mugilidae	*Chelon labrosus*	1.23%
Nemacheilidae	***Barbatula barbatula***	82.71%
Percidae	*Gymnocephalus cernuus*	25.92%
Percidae	*Perca fluviatilis*	53.8%
Percidae	*Sander lucioperca*	19.75%
Petromizontidae	*Lampetra fluviatilis*	1.23%
Petromizontidae	*Lampetra planeri*	43.20%
Petromizontidae	*Petromyzon marinus*	11.11%

Pleuronectidae	*Platichthys flesus*	1.23%
Poeciliidae	*Gambusia holbrooki*	4.93%
Salmonidae	*Oncorhynchus mykiss*	16.4%
Salmonidae	*Salmo salar*	8.64%
Salmonidae	*Salmo trutta*	69.13%
Salmonidae	*Salvelinus fontinalis*	6.17%
Salmonidae	*Thymallus thymallus*	4.93%
Siluridae	*Silurus glanis*	22.22%
Umbridae	*Umbra pygmaea*	1.23%

Table II-S3: Mean effect size ($\bar{E}$), 95% confidence interval (95% CI), total heterogeneity of a sample (Q_t) and corresponding p-value (P) computed from the path coefficients obtained from partial least-square path modeling applied to (a) α-diversity indices and (b) β-diversity indices computed after the removal of non-native species. Values in bold are significant mean effect sizes and the corresponding 95% confidence intervals.

(a) **α-diversity**	$\bar{E}$	95% CI	Q_t	P
Isolation → Allelic richness	**-0.463**	**[-0.683; -0.248]**	19.861	0.001
Area → Allelic richness	**0.237**	**[0.017; -0.458]**	6.669	0.083
Connectivity → Allelic richness	0.079	[-0.141; 0.299]	1.778	0.619
Water composition → Allelic richness	0.121	[-0.099; 0.341]	3.328	0.343
Oxygen → Allelic richness	0.175	[-0.045; 0.395]	3.652	0.301
Isolation → Species richness	**-0.354**	**[-0.574; -0.134]**	1.481	0.686
Area → Species richness	**0.404**	**[0.184; 0.624]**	2.934	0.401
Connectivity → Species richness	-0.068	[-0.288; 0.151]	0.976	0.807
Water composition → Species richness	0.131	[-0.089; 0.351]	9.053	0.028
Oxygen → Species richness	-0.214	[-0.434; 0.007]	0.251	0.975
Allelic richness → Species richness	0.006	[-0.214; 0.226]	0.209	0.976
(b) **β-diversity**	$\bar{E}$	95% CI	Q_t	P
Isolation → Jost's D	**0.362**	**[0.322; 0.402]**	250.673	0
Area → Jost's D	**-0.053**	**[-0.093; -0.013]**	22.164	0
Connectivity → Jost's D	0.006	[-0.034; 0.046]	44.912	0
Water composition → Jost's D	**0.126**	**[0.086; 0.166]**	282.503	0
Oxygen → Jost's D	-0.018	[-0.058; 0.022]	35.124	0
Isolation → True diversity	0.012	[-0.028; 0.051]	17.941	0
Area → True diversity	-0.016	[-0.056; 0.024]	7.539	0.057
Connectivity → True diversity	0.015	[-0.255; 0.054]	16.746	0.001
Water composition → True diversity	**0.143**	**[0.103; 0.183]**	605.119	0
Oxygen → True diversity	**-0.053**	**[-0.093; -0.013]**	15.149	0.002
Jost's D → True diversity	**0.080**	**[0.040; 0.120]**	10.250	0.017

Table II-S4: Sampling information on microsatellite loci and multiplexed PCR used in this study for the four species.

Species	Locus	GenBank Accession ID	Fluorescent Dye	Multiplex Kit	Primer Concentration (nM)	Reference
Squalius cephalus	Ca01	AF277573	HEX	1	100	Dimsoski et al. 2000
	CypG30	AY439148	NED	1	150	Baerwald et al. 2004
	Lc293	EF362795	FAM	1	100	Vyskocilova et al. 2007
	Lcel100	AY962249	FAM	1	300	Larno et al. 2005
	Lc290	EF362794	HEX	1	100	Vyskocilova et al. 2007
	CypG24	AY439142	FAM	2	50	Baerwald et al. 2004
	CypG27	AY439145	HEX	2	200	Baerwald et al. 2004
	Lc27	EF362792	NED	2	100	Vyskocilova et al. 2007
	LceC1	AY962241	FAM	2	100	Larno et al. 2005
	Lid11	AB112736	FAM	2	250	Barinova et al. 2004
Gobio occitaniae	Ca01	AF277573	HEX	1	100	Dimsoski et al. 2000
	Gob16	DQ207803	HEX	1	30	Knapen et al. 2006
	Rhca20	DQ106915	NED	1	100	Girard and Angers 2006
	Gob12	DQ207801	FAM	1	175	Knapen et al. 2006
	Gob15	DQ207802	HEX	2	100	Knapen et al. 2006
	Gob22	DQ207804	FAM	2	50	Knapen et al. 2006
	Gob28	DQ207805	HEX	2	100	Knapen et al. 2006
	MFW1	n.a.	NED	2	100	Crooijmans et al. 1997
Barbatula barbatula	Bbar1	AF311346	NED	1	40	Taylor et al. 2001
	Bbar7	AF310881	FAM	1	80	Taylor et al. 2001
	Bbar8	AF310882	HEX	1	200	Taylor et al. 2001
	Lec12	AB286043	NED	1	80	Koizumi et al. 2007
	Bbar11	AF310883	HEX	2	80	Taylor et al. 2001
	Bbar9	AF311349	NED	2	60	Taylor et al. 2001
	Lec01	AB286032	FAM	2	80	Koizumi et al. 2007
	Lec05	AB286036	HEX	2	150	Koizumi et al. 2007
Phoxinus phoxinus	CypG27	AY439145	HEX	1	200	Baerwald et al. 2004
	CypG30	AY439148	NED	1	100	Baerwald et al. 2004
	LceC1	AY962241	FAM	1	50	Larno et al. 2005
	Lid2	AB112733	FAM	1	200	Barinova et al. 2004
	Rru4	AB112740	HEX	1	20	Barinova et al. 2004
	Ca12	AF277584	FAM	2	130	Dimsoski et al. 2000
	CypG03	AY439122	HEX	2	50	Baerwald et al. 2004
	Lc27	EF362792	NED	2	100	Vyskocilova et al. 2007
	MFW1	n.a.	HEX	2	250	Crooijmans et al. 1997
	Rhca20	DQ106915	FAM	2	50	Girard and Angers 2006

References for Table II-S4:

Baerwald M. R. & May B. 2004. Characterization of microsatellite loci for five members of the minnow family Cyprinidae found in the Sacramento-San Joaquin Delta and its tributaries. *Molecular Ecology Notes* **4**, 385-390.

Barinova A., Yadrenkina E., Nakajima M. & Taniguchi N. 2004. Identification and characterization of microsatellite DNA markers developed in ide *Leuciscus idus* and Siberian roach *Rutilus rutilus*. *Molecular Ecology Notes* **4**, 86-88.

Crooijmans R., Bierbooms V.A.F, Komen J., VanderPoel J.J. & Groenen M.A.M. 1997. Microsatellite markers in common carp (*Cyprinus carpio* L). *Animal Genetics* **28**, 129-134.

Dimsoski P., Toth G.P. & Bagley M.J. 2000. Microsatellite characterization in central stoneroller *Campostoma anomalum* (Pisces : Cyprinidae). *Molecular Ecology* **9**, 2187-2189.

Girard P. & Angers B. 2006. Characterization of microsatellite loci in longnose dace (*Rhinichthys cataractae*) and interspecific amplification in five other Leuciscinae species. *Molecular Ecology Notes* **6**, 69-71.

Koizumi N., Jinguji H., Takahashi H., *et al.* 2007 Isolation and characterization of polymorphic microsatellite DNA markers in the Omono type of ninespine stickleback, genus *Pungitius*. *Molecular Ecology Notes* **7**, 1315–1318.

Knapen D., Taylor M.I., Blust R. & Verheyen E. 2006. Isolation and characterization of polymorphic microsatellite loci in the gudgeon, *Gobio gobio* (Cyprinidae). *Molecular Ecology Notes* **6**, 387-389.

Larno V., Launey S., Devaux A. & Laroche J. 2005. Isolation and characterization of microsatellite loci from chub *Leuciscus cephalus* (Pisces : Cyprinidae). *Molecular Ecology Notes* **5**, 752-754.

Taylor M.I., Blust R., and Verheyen E. 2001. Characterization of microsatellite loci in the stone loach, *Barbatula barbatula L. Molecular Ecology Notes* **1**, 96–97.

Turner T.F., Dowling T.E., Broughton R.E., and J.R. Gold. 2004. Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily *Leusicinae*). *Conservation Genetics* **5**, 279-281.

Vyskocilova M., Simkova A., and Martin J.F. 2007. Isolation and characterization of microsatellites in *Leuciscus cephalus* (Cypriniformes, Cyprinidae) and cross-species amplification within the family Cyprinidae. *Molecular Ecology Notes* **7**, 1150-1154.

Table II-S5: Correlation values between the latent variables of the eight models.

(a) At the α-level:

Barbatula barbatula	Isolation	Area	Connectivity	Water composition	Oxygen	Allelic richness	Species richness
Isolation	1.0000	-0.0639	-0.1011	-0.5506	0.1198	-0.454	-0.490
Area	-0.0639	1.0000	0.5767	-0.3147	-0.0234	0.348	0.291
Connectivity	-0.1011	0.5767	1.0000	-0.0194	0.1092	0.327	0.154
Water composition	-0.5506	-0.3147	-0.0194	1.0000	-0.1111	0.322	0.409
Oxygen	0.1198	-0.0234	0.1092	-0.1111	1.0000	0.169	-0.327
Allelic richness	-0.4542	0.3483	0.3273	0.3216	0.1686	1.000	0.278
Species richness	-0.4905	0.2908	0.1538	0.4093	-0.3269	0.278	1.000

Gobio occitaniae	Isolation	Area	Connectivity	Water composition	Oxygen	Allelic richness	Species richness
Isolation	1.000	-0.407	-0.345	-0.643	0.268	-0.645	-0.576
Area	-0.407	1.000	0.649	0.421	-0.110	0.252	0.505
Connectivity	-0.345	0.649	1.000	0.459	-0.126	0.198	0.355
Water composition	-0.643	0.421	0.459	1.000	-0.224	0.410	0.525
Oxygen	0.268	-0.110	-0.126	-0.224	1.000	0.125	-0.360
Allelic richness	-0.645	0.252	0.198	0.410	0.125	1.000	0.280
Species richness	-0.576	0.505	0.355	0.525	-0.360	0.280	1.000

Phoxinus phoxinus	Isolation	Area	Connectivity	Water composition	Oxygen	Allelic richness	Species richness
Isolation	1.0000	-0.2090	-0.2097	-0.5202	0.0939	-0.4470	-0.539
Area	-0.2090	1.0000	0.6108	-0.0176	0.0030	0.3623	0.445
Connectivity	-0.2097	0.6108	1.0000	0.2024	0.0871	0.3508	0.228
Water composition	-0.5202	-0.0176	0.2024	1.0000	0.0143	0.3507	0.417
Oxygen	0.0939	0.0030	0.0871	0.0143	1.0000	0.0634	-0.223
Allelic richness	-0.4470	0.3623	0.3508	0.3507	0.0634	1.0000	0.331
Species richness	-0.5389	0.4448	0.2283	0.4168-	0.2226	0.3313	1.000

Squalus cephalus	Isolation	Area	Connectivity	Water composition	Oxygen	Allelic richness	Species richness
Isolation	1.000	-0.3643	-0.316	-0.686	0.2264	-0.4034	-0.541
Area	-0.364	1.0000	0.668	0.605	-0.0815	0.5826	0.465
Connectivity	-0.316	0.6676	1.000	0.596	-0.1087	0.5225	0.360
Water composition	-0.686	0.6048	0.596	1.000	-0.2828	0.4914	0.523
Oxygen	0.226	-0.0815	-0.109	-0.283	1.0000	-0.0801	-0.230
Allelic richness	-0.403	0.5826	0.522	0.491	-0.0801	1.0000	0.405
Species richness	-0.541	0.4649	0.360	0.523	-0.2304	0.4053	1.000

(b) At the β-level:

Barbatula barbatula	Isolation	Area	Connectivity	Water composition	Oxygen	Jost'sD	True diversity
Isolation	1.0000	-0.0458	-0.0884	-0.551	0.1138	0.3854	0.429
Area	-0.0458	1.0000	0.5767	-0.317	-0.0264	0.1364	-0.188
Connectivity	-0.0884	0.5767	1.0000	-0.029	0.1052	-0.0655	-0.245
Water composition	-0.5506	-0.3173	-0.0290	1.000	-0.1178	-0.3040	-0.332
Oxygen	0.1138	-0.0264	0.1052	-0.118	1.0000	-0.1439	0.200
Jost'sD	0.3854	0.1364	-0.0655	-0.304	-0.1439	1.0000	0.124
True diversity	0.4290	-0.1882	-0.2455	-0.332	0.2002	0.1238	1.000

Gobio occitaniae	Isolation	Area	Connectivity	Water composition	Oxygen	Jost'sD	True diversity
Isolation	1.000	-0.4112	-0.3513	-0.169	0.2596	0.0720	0.2382
Area	-0.411	1.0000	0.6489	0.240	-0.0983	-0.1294	-0.0235
Connectivity	-0.351	0.6489	1.0000	0.277	-0.1165	-0.0404	-0.0028
Water composition	-0.169	0.2404	0.2773	1.000	0.1019	-0.2166	-0.3418
Oxygen	0.260	-0.0983	-0.1165	0.102	1.0000	-0.3213	0.1013
Jost'sD	0.072	-0.1294	-0.0404	-0.217	-0.3213	1.0000	0.1575
True diversity	0.238	-0.0235	-0.0028	-0.342	0.1013	0.1575	1.0000

Phoxinus phoxinus	Isolation	Area	Connectivity	Water composition	Oxygen	Jost'sD	True diversity
Isolation	1.000	-0.2480	-0.2402	-0.5050	0.1364	0.1317	0.4465
Area	-0.248	1.0000	0.6108	0.0804	-0.0243	-0.3340	-0.2842
Connectivity	-0.240	0.6108	1.0000	0.2827	0.0527	-0.2910	-0.2215
Water composition	-0.505	0.0804	0.2827	1.0000	-0.0169	-0.2650	-0.4867
Oxygen	0.136	-0.0243	0.0527	-0.0169	1.0000	0.1557	0.1631
Jost'sD	0.132	-0.3340	-0.2910	-0.2650	0.1557	1.0000	0.0733
True diversity	0.447	-0.2842	-0.2215	-0.4867	0.1631	0.0733	1.0000

Squalus cephalus	Isolation	Area	Connectivity	Water composition	Oxygen	Jost'sD	True diversity
Isolation	1.0000	-0.3630	-0.3142	0.674	0.2321	0.5614	0.0338
Area	-0.3630	1.0000	0.6676	-0.326	-0.0904	-0.2607	0.0682
Connectivity	-0.3142	0.6676	1.0000	-0.295	-0.1151	-0.0396	0.1201
Water composition	0.6738	-0.3258	-0.2949	1.000	0.2759	0.3786	-0.3209
Oxygen	0.2321	-0.0904	-0.1151	0.276	1.0000	0.0163	-0.0200
Jost'sD	0.5614	-0.2607	-0.0396	0.379	0.0163	1.0000	0.1376
True diversity	0.0338	0.0682	0.1201	-0.321	-0.0200	0.1376	1.0000

Table II-S6: Tables for Hardy-Weinberg and linkage disequilibrium tests for all species. Table II-S6 is available online at: http://onlinelibrary.wiley.com/doi/10.1111/fwb.12826/abstract

Table II-S7: Fisher's Z effect sizes of each species and each path computed from the path coefficients obtained from partial least-square path modeling applied to (a) α-diversity indices and (b) β-diversity indices. Values in bold are Fisher's Z effect sizes leading to significant mean effect sizes ($\bar{E}$, see Table II-2).

(a) **α-diversity**	*Barbatula barbatula*	*Gobio occitaniae*	*Phoxinus phoxinus*	*Squalius cephalus*
Isolation → Allelic richness	**-0.269**	**-0.721**	**-0.286**	**-0.198**
Area → Allelic richness	**0.420**	**-0.004**	**0.247**	**0.361**
Connectivity → Allelic richness	0.038	-0.022	0.095	0.211
Water composition → Allelic richness	0.333	0.031	0.186	0.017
Oxygen → Allelic richness	0.243	0.321	0.079	0.022
Isolation → Species richness	**-0.240**	**-0.330**	**-0.298**	**-0.347**
Area → Species richness	**0.482**	**0.334**	**0.511**	**0.244**
Connectivity → Species richness	-0.094	-0.074	-0.187	-0.012
Water composition → Species richness	**0.424**	**0.185**	**0.319**	**0.072**
Oxygen → Species richness	-0.218	-0.196	-0.183	-0.106
Allelic richness → Species richness	-0.068	-0.054	-0.021	0.085
(b) **β-diversity**	*Barbatula barbatula*	*Gobio occitaniae*	*Phoxinus phoxinus*	*Squalius cephalus*
Isolation → Jost's D	**0.536**	**0.035**	**0.380**	**0.484**
Area → Jost's D	0.074	-0.062	0.021	-0.052
Connectivity → Jost's D	-0.053	0.030	-0.126	0.151
Water composition → Jost's D	**-0.023**	**0.394**	**0.042**	**0.022**
Oxygen → Jost's D	-0.037	0.071	-0.129	-0.048
Isolation → True diversity	-0.029	-0.040	0.126	-0.073
Area → True diversity	-0.108	-0.026	-0.051	-0.004
Connectivity → True diversity	**-0.026**	**0.073**	**0.042**	**0.067**
Water composition → True diversity	**0.104**	**0.247**	**0.140**	**0.334**
Oxygen → True diversity	**-0.086**	**-0.126**	**-0.056**	**-0.127**
Jost's D → True diversity	**0.181**	**0.103**	**0.037**	**0.058**

Chapter III - Genetic-phenotypic intraspecific diversity covariation in structured landscapes: An empirical test of underlying determinants

By Lisa Fourtune[1], Jérôme G. Prunier[1], Églantine Mathieu-Bégné[1, 2, 3], Nicolas Canto[1], Charlotte Veyssière[2, 3], Géraldine Loot[2, 3] & Simon Blanchet[1,3]

[1]Centre National de la Recherche Scientifique (CNRS), Université Sabatier (UPS), UMR 5321 (Station d'Écologie Théorique et Expérimentale), Moulis, France

[2]Université de Toulouse, UPS, UMR 5174 (Laboratoire Évolution & Diversité Biologique, EDB), Toulouse, France

[3]CNRS, UPS, École Nationale de Formation Agronomique (ENFA), UMR 5174 EDB, Toulouse, France

III.1 - Résumé

La diversité intraspécifique joue un rôle majeur au cœur des dynamiques évolutive et écologique. Elle est le matériau de base sur lequel agit la sélection, elle améliore la résilience des espèces et des communautés face aux perturbations et elle influence la manière dont les espèces modifient leurs environnements biotique et abiotique. Comprendre les patrons spatiaux de diversité intraspécifique ainsi que les processus qui les façonnent est donc crucial. Nous nous sommes ici intéressés à deux poissons d'eau douce (*Gobio occitaniae* and *Phoxinus phoxinus*) que nous avons échantionnés au sein du bassin versant de la Garonne-Dordogne afin de rechercher l'existence de corrélations entre diversité génétique et diversité phénotypique chez ces deux espèces. Nous avons également étudié les processus déterminant la distribution de ces deux facettes de la diversité intraspécifique par l'utilisation d'analyses causales.La diversité génétique a été estimée à partir de marqueurs microsatellites, et la diversité phénotypique a été mesurée par des analyses de morphométrie géométrique. Nous avons mis en évidence des disparités entre les patrons de diversité génétique et phénotypique de ces deux espèces semblant indiquer une adaptation locale plus forte chez *G. occitaniae* et nos résultats ont révélé que des processus communs et distincts façonnaient ces patrons. Au niveau α, nous n'avons trouvé aucune corrélation entre diversité génétique et diversité phénotypique, malgré des relations similaires entre isolation et diversité génétique et phénotypique chez *G. occitaniae*. Au niveau β, nous n'avons pas trouvé de corrélation entre diversité génétique et diversité phénotypique chez *P. phoxinus*, mais nous en avons mise une en évidence chez *G. occitaniae.* Cette corrélation est apparue comme provenant d'un impact direct d'une des deux facettes de diversité intraspécifique sur l'autre, et nous avons émis l'hypothèse qu'elle pourrait résulter d'appariements sélectifs chez cette espèce. L'étude de la diversité génétique et phénotypique par une approche intégrative semble être une méthode précieuse pour analyser les nombreux impacts des processus neutres et adaptatifs sur les patrons de diversité intraspécifique.

III.2 - Abstract

Intraspecific diversity plays a key role for evolutionary and ecological dynamics. It is the raw material on which acts selection, it improves species and communities resilience to disturbance and it affects the way species modulate their biotic and abiotic environment. Understanding patterns and underlying determinants of genetic and phenotypic intraspecific diversity is therefore of critical importance for ecological, evolutionary and conservation sciences. Here, focusing on two freshwater fish species (*Gobio occitaniae* and *Phoxinus phoxinus*) sampled across a large river basin (the Garonne-Dordogne river basin, France), we used causal analyses to test for genetic-phenotypic intraspecific diversity correlations (GPIDCs) and unravel the processes underlying intraspecific diversity patterns. Genetic diversity was assessed using microsatellite markers and phenotypic diversity was assessed through geometric morphometrics. We found disparities in the distribution of genetic and phenotypic diversity in the two species, suggesting higher level of local adaptation in *G. occitaniae*, and our results revealed common and contrasted processes shaping diversity at the α- and β-level. At the α-level, we found no GPIDC in both species despite common relations between isolation and genetic and phenotypic α-diversity in *G. occitaniae*. At the β-level, we found no GPIDC in *P. phoxinus* but we found a positive GPIDC in *G. occitaniae*. This correlation appeared to be caused by a direct impact of one facet of intraspecific diversity on the other, and we speculated that it could originate from positive assortative mating. Studying neutral genetic diversity and phenotypic diversity within an integrative framework appears as a valuable way of deciphering the complex and diverse impacts of neutral and adaptive processes on intraspecific diversity patterns.

III.3 - Introduction

Within non-clonal species, all individuals are genetically and phenotypically unique, which constitutes the most elemental facet of biological diversity. Intraspecific biodiversity plays a key role for evolutionary and ecological dynamics (Bolnick *et al.* 2003; Odling-Smee, Laland & Feldman 2003). It is the raw material on which acts selection, potentially leading to adaptation to environmental changes, and improving species and communities resilience to disturbances (Jung *et al.* 2013; Moran, Hartig & Bell 2015). Intraspecific diversity also affects the way species modulate their biotic and abiotic environment, thus impacting community structure and ecosystem functioning (Hughes *et al.* 2008; Bolnick *et al.* 2011). Therefore, understanding patterns and underlying determinants of intraspecific diversity is of critical importance for ecological, evolutionary and conservation sciences (Chave 2013; Mimura *et al.* 2017).

By analogy to interspecific diversity, intraspecific diversity can be decomposed into two components: within-population (intraspecific α-diversity) and between-population intraspecific diversity (intraspecific β-diversity) (Loreau 2000). Within-population intraspecific diversity corresponds to the diversity space covered by the individuals of a given population, whereas between-population intraspecific diversity corresponds to the differentiation observed among populations. Intraspecific diversity also comprises a genetic and a phenotypic (here phenotypes include behavioural, morphological and physiological traits) facet, the former being inherited from the parents and the later being affected by both inherited and non-inherited (environmental) mechanisms. Intraspecific genetic diversity is defined as the variability of neutral and adaptive genetic sequences observed within populations (Holderegger, Kamm & Gugerli 2006), whereas phenotypic diversity encompasses the diversity of individuals' traits.

Understanding how intraspecific diversity is maintained at the population level has attracted both ecologists and evolutionists for decades. For instance, a surge of studies have focused on describing patterns of intraspecific neutral genetic diversity (e.g. through allelic richness and F_{ST}), so as to unravel the demographic and evolutionary history of populations, and hence to improve their conservation and management (Manel *et al.* 2003; Reed & Frankham 2003; Blanchet, Prunier & De Kort 2017). From an adaptive point of view, the relative importance of divergent natural selection in shaping the distribution of non-neutral traits across landscapes -and hence phenotypic β-diversity- has been the focus of both

quantitative genetics and experimental approaches (Kawecki & Ebert 2004; Leinonen *et al.* 2013; Blanquart *et al.* 2013). In parallel, ecologists have recently focused on the distribution of intraspecific phenotypic α-diversity across species and landscapes in order to better appraise its roles for ecosystem functioning and community dynamics (Violle *et al.* 2012; Moran *et al.* 2015; Siefert *et al.* 2015). However, the study of intraspecific diversity still lacks an integrative framework in which patterns of genetic and phenotypic diversity, as well as their underlying determinants, would be investigated simultaneously and considered as two potentially covarying facets of biological diversity. Remarkably, a framework in which two facets of biodiversity (namely species diversity and intraspecific genetic diversity) are studied integratively has been introduced by Vellend (2005) and has generated an increasing number of studies (reviewed in Vellend *et al.* 2014; Lamy *et al.* 2017). These studies on species-genetic diversity correlations (SGDCs) led to a better understanding of the relationships between species and genetic diversity, as well as the processes shaping these facets of biodiversity in similar or contrasting ways (Taberlet *et al.* 2012; Vellend *et al.* 2014).

Studying genetic-phenotypic intraspecific diversity correlations (GPIDCs) within an analogous framework appears reasonable as these two facets of diversity are intrinsically related and can be under the influence of similar adaptive and neutral processes (Lowe, Kovach & Allendorf 2017). For instance, in the case of non-neutral genetic markers and adaptive traits, a positive GPIDC is expected when genetic and phenotypic non-neutral diversity are directly affected by environmental conditions (through selection and/or plasticity). In this case, genetic and phenotypic α-diversity are expected to be high in populations inhabiting highly heterogeneous environments, and genetic and phenotypic β-diversity are expected to be high between populations experiencing contrasting environmental conditions (Leimar 2005; Hedrick 2006; Wang & Bradburd 2014). Genetic and phenotypic diversity are also expected to be positively correlated if they are driven by neutral processes such as drift and dispersal (but see Edelaar, Siepielski & Clobert 2008; Lowe & McPeek 2014), which can notably be the case for neutral genetic markers and phenotypic traits that are not strongly affected by selection (Hartl & Clark 2007). Under this hypothesis, genetic and phenotypic α-diversity should be high in populations with large effective sizes and/or experiencing strong immigration. At the β-level, genetic and phenotypic diversity should be high between populations undergoing strong genetic drift (Prunier *et al.* 2017) and/or geographically isolated one from each other (Hutchison & Templeton 1999). Finally, a positive GPIDC can be explained by a direct relationship between genetic and phenotypic

diversity, notably when genetic diversity directly codes for the considered traits or appropriately describes the whole genomic diversity (Hoffman *et al.* 2014). Conversely, when genetic and phenotypic diversity are driven by divergent processes, GPIDCs are expected to be non-significant.

Here, we aimed at testing for spatial covariations in genetic and phenotypic intraspecific diversity in two sympatric species inhabiting a spatially-structured landscape, and at unravelling underlying determinants at a large spatial scale. More specifically, we first quantified and described genetic and phenotypic intraspecific diversity in two sympatric freshwater fish species (*Gobio occitaniae* and *Phoxinus phoxinus*) across an entire river drainage. We then investigated both α- and β-GPIDCs for these two species, and we finally deciphered the parallel or independent determinants shaping genetic and phenotypic diversity at the α- and β-levels using causal analyses. To this end, we gathered neutral genetic diversity and morphological diversity in both *G. occitaniae* and *P. phoxinus* so as to test whether or not the relative importance of main determinants of GPIDCs varied for species sharing a similar environment but with different life-history traits. We predicted that GPIDCs should be weak for the two species since neutral genetic diversity should mainly be driven by gene flow and/or drift, whereas morphology should be determined by environmental characteristics. Alternatively, in dendritic ecological networks (Grant, Lowe & Fagan 2007) such as riverine networks, we can predict positive GPIDCs because factors affecting neutral processes (e.g. habitat areas) and adaptive processes (e.g. physico-chemical conditions) tend to covary along the network. Moreover, their specific structure (treelike branching, constrained dispersal, upstream-downstream environmental gradient) has already been theoretically and empirically shown to affect patterns of neutral and non-neutral diversity (Paz-Vinas & Blanchet 2015; Fronhofer & Altermatt 2017). Testing GPIDCs in highly spatially-structured landscapes such as dendritic riverine networks thus appears of particular interest.

III.4 - Materials and Methods

Collection of genetic and phenotypic data

Study species. *Gobio occitaniae* (the Occitan gudgeon) and *Phoxinus phoxinus* (the European minnow) belong to the Cyprinidae family. Both species are insectivorous but differ in their foraging mode: *G. occitaniae* feeds predominantly on the bottom, whereas *P. phoxinus*

feeds in the water column. *Gobio occitaniae* mean body length (120-150mm) is slightly larger than that of *P. phoxinus* (80-90mm). Moreover these species have contrasting levels of habitat specialisation: G. occitaniae lives in many habitat types and is found almost everywhere in the river basin whereas P. phoxinus lives preferentially in upstream sections.

Study area and sampling. Fish were sampled across 48 sites evenly scattered across the Garonne-Dordogne river drainage (South-Western France). This river drainage covers a 79 800km² area and sites were selected so as to cover the whole distribution of the two fish species, and hence their entire realized environmental niches. Electrofishing sampling was conducted during summers 2014 (42 sites) and 2015 (6 sites) and each site was visited once. Sampled area was of ~500-1000 m2 to adequately represent the local habitat heterogeneity. *Gobio occitaniae* and *P. phoxinus* individuals were found in 39 and 34 sites respectively, with 25 sites in which the species were sympatric (see Figure III-1). We sampled up to thirty individuals per species and per sampling site (range, 21-30 and 24-30 for *G. occitaniae* and *P. phoxinus* respectively, see Table III-S1), leading to 1119 gudgeons and 978 minnows. Sampled individuals were anaesthetised using oil of clove before being carefully aligned on their right side on a white dashboard including a reference scale. The left side of each individual was photographed using a digital camera (Canon G16©) mounted on a tripod. Subsequently, we collected on each individual a small piece of pelvic fin which was preserved in 70% ethanol for genetic analyses. Individuals were then released alive in their respective sampling site.

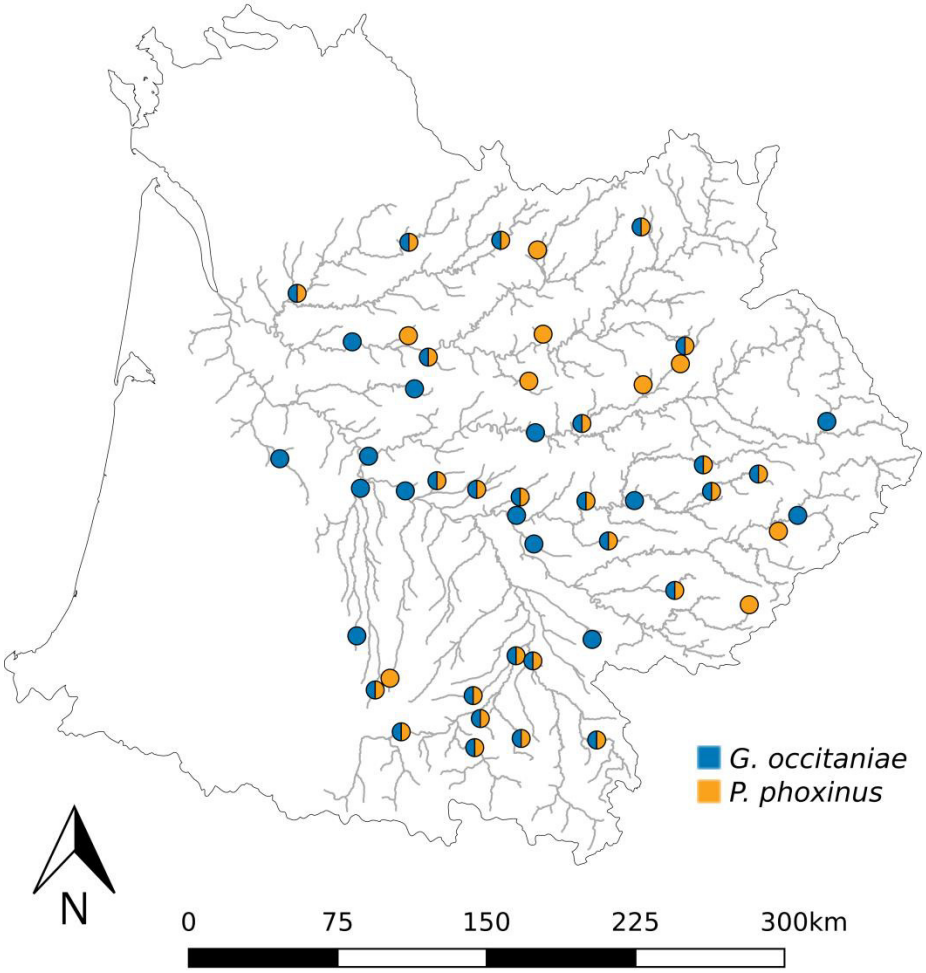


Figure III-1: Location of the 48 sites sampled during summers 2014 and 2015 colored according to the species present.

Genetic data. Genetic DNA was extracted from all samples using a salt-extraction protocol (Aljanabi & Martinez 1997). Genotyping was performed using 15 and 18 microsatellite loci in *G. occitaniae* and *P. phoxinus* respectively. Polymerase chain reactions (PCR) were performed as in (Blanchet *et al.* 2010). Genotypes were analysed using GENEMAPPER 5.0 (Applied Biosystems©). The presence of null alleles was assessed at each locus using MICROCHECKER 2.2.3 (Van Oosterhout *et al.* 2004). We also checked for gametic disequilibrium using GENEPOP 4.2.1 (Rousset 2008) after sequential Bonferroni correction to account for multiple tests. We discarded from further analyses any locus showing significant gametic disequilibrium and/or evidence of null alleles, leading to a total of 13 and 17 loci for *G. occitaniae* and *P. phoxinus* respectively.

As a measure of genetic α-diversity, we computed -for each species- the allelic

richness (i.e. the mean number of alleles across loci standardized for the lowest sample size, i.e. n = 20 for the two species) using ADZE 1.0 (Szpiech, Jakobsson & Rosenberg 2008). As a measure of genetic β-diversity, we used three common indices based on allelic frequencies: Rousset's linearized F_{ST} ($F_{ST}/(1-F_{ST})$), hereafter denoted as F_{ST} (Rousset 1997), Nei's version of Cavalli-Sforza's chord distance *Da* (Nei, Tajima & Tateno 1983) and Jost's D (Jost 2008). Whatever the dataset, these three indices were highly correlated (Mantel $r > 0.85$, $p < 0.001$): for the sake of simplicity, we thus only retained F_{ST} as a measure of genetic differentiation.

Phenotypic data. Individuals morphology was analysed using a landmark-based geometric morphometrics approach (Rohlf & Marcus 1993). Sixteen homologous landmarks were defined so as to capture the overall body shape of each individual (Figur III-2). Landmarks coordinates were obtained from digitized pictures using the Pointpicker plugin (http://bigwww.epfl.ch/thevenaz/pointpicker/) in the ImageJ software (Schneider, Rasband & Eliceiri 2012). As the distance between the camera and the fish slightly varied between sites, we size-corrected landmarks coordinates using the reference scale. For each species, landmarks were aligned using Generalized Procrustes Analysis (Rohlf & Slice 1990) with the R package *geomorph* (Adams & Otárola-Castillo 2013) in order to remove the effects of rotation, translation and scale on shape variation. Relative warps (n = 32) were computed for each individual by performing a Principal Component Analysis (PCA) on the aligned landmark coordinates of each species (Rohlf 1993). As the majority of the 32 relative warps explained a very small amount of variation, we only conserved for further analyses the first nine relative warps that together explained more than 85% of the variance in each data set (85.69% and 85.32% for *G. occitaniae* and *P. phoxinus* respectively). These relative warps were used as shape variables. Individual centroid size, which is the square root of the sum of squared distances from landmarks to their centroid, was used as a surrogate of overall body size of each individual (Bookstein 1991).

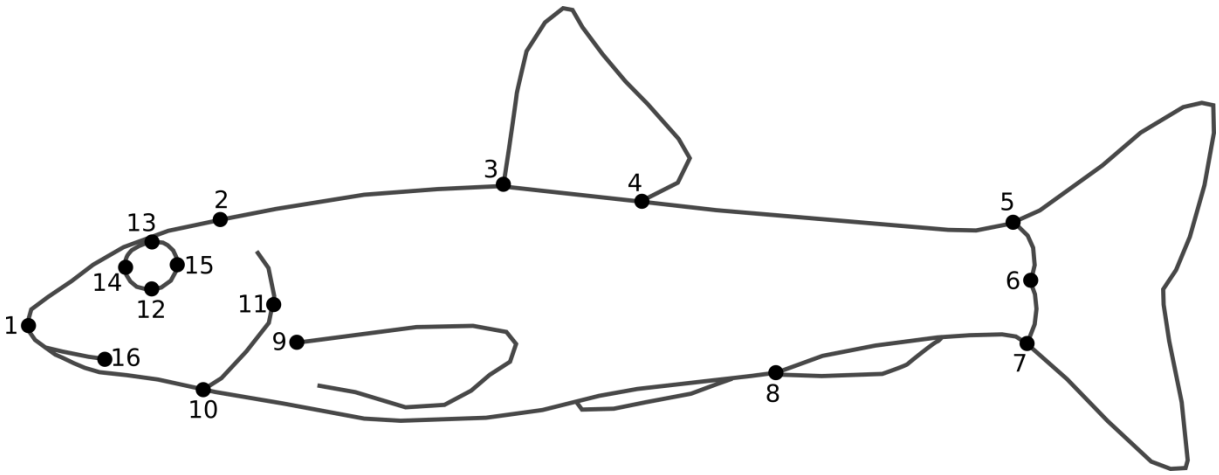


Figure III-2: Location of 16 homologous landmarks used to assess phenotypic diversity in *Gobio occitaniae* and *Phoxinus phoxinus*. Landmarks refer to (1) tip of the snout, (2) beginning of scales coverage on the dorsal outline, (3) anterior and (4) posterior insertions of the dorsal fin, (5) dorsal insertion of the caudal fin, (6) posterior extremity of the body, (7) ventral insertion of the caudal fin, (8) anterior insertion of the anal fin, (9) superior insertion of the pectoral fin, (10) posterior border of the operculum, (11) posterior extremity of the operculum, (12) the inferior, (13) superior, (14) anterior and (15) posterior extremities of the orbital circumference, (16) posterior extremity of the premaxillar.

Since relative warps are PCA coordinates, they can be seen as coordinates in a nine dimensions shape space (i.e. the hyperspace in which each point represents a configuration of landmarks) of each individual. Morphological α-diversity was thus computed as the proportion of the total shape space (observed over all populations for a given species) occupied by the individuals of a population (see Figure III-S2), after accounting for differences in sampling sizes among populations using a random resampling approach. This index is equivalent to the functional richness index developed by (Villéger, Mason & Mouillot 2008) at the interspecific level. Morphological β-diversity was computed -for each species- as the euclidean distance between the consensus (i.e. the average shape in a population) of each population pair (see Figure III-S2). Additionally, in order to further decompose this index of differentiation, we computed the functional dissimilarity index F_β developed by (Villéger, Novack-Gottshall & Mouillot 2011) at the interspecific level and informing the proportion of the total shape space that is not shared between two populations from a given pair:

$$F_\beta = 1 - (\text{Volume shared})/(\text{Total volume})$$

F_β ranges from 0 to 1, with 0 indicating a perfect overlap in shape space occupation and 1 indicating that no space is shared between two populations. Intermediate F_β can result from two (non-exclusive) mechanisms: turnover (the two populations fill distinct parts of the shape space with weak overlap) and nestedness (one population fills a small proportion of the shape space filled by the other) (Baselga 2010; Villéger, Grenouillet & Brosse 2013). Consequently,

we further computed F_{pturn}, the proportion of F_{β} that is due to trait turnover so as to tease apart the effect of nestedness and turnover. Because of computing limitations, these two indices were computed only from the first three relative warps coordinates. All indices were computed using functions available at http://villeger.sebastien.free.fr/Rscripts.html.

Collection of environmental data

We gathered several variables related to environmental characteristics and river topography. These variables are likely to impact intraspecific diversity through evolutive and/or neutral processes. General expectations are listed in Table III-1.

Environmental characteristics. Substrate specificity was evaluated visually on each site following a predefined protocol: substrate was classified into nine categories based on particle size, ranging from silt (< 0.05mm) to solid bedrock (see Table III-S3), and the percentage of each category composing the river bed of each site was estimated visually within a predefined area of ~100 m^2. From these data, habitat heterogeneity was computed as Pielou's evenness index, with low values identifying sites in which one of the substrate categories was dominant (i.e. sites of homogeneous substrate). Habitat dissimilarities between sites were computed from habitat percentages as Bray-Curtis distances, with a value of 1 identifying two sites sharing no substrate categories. The other environmental variables were obtained for each site from the database of the Water Information System of the Adour Garonne basin (SIEAG, 'Système d'Information sur l'Eau du Bassin Adour Garonne'; http://adour-garonne.eaufrance.fr) that gathers physico-chemical characteristics of surface water measured several times every year at numerous sites in the river catchment. Only sites for which data were available for July (a month in which the two species are highly active) of the years 2013, 2014 and 2015 were selected from the SIEAG database. The mean of the three values was calculated to inform the physico-chemical quality of the sites according to several parameters. We notably focused on parameters directly affecting fish populations, i.e. oxygen saturation (%) (Crispo & Chapman 2008) and water temperature (°C) (Buisson, Blanc & Grenouillet 2008). Moreover, we gathered eleven variables informing water quality: concentrations in ammonium, azote, organic carbon, nitrate, nitrite, orthophosphate and phosphorus (mg/L), Biological Oxygen Demand (mg/L), water conductivity (mS/cm), pH and suspended matter (mg/L). We performed a PCA on these variables, and gathered the coordinates of each site on the first two axes (representing respectively 36.85% and 19.73% of the total variance) to create two synthetic variables (hereafter named chemicals1 and

chemicals2) informing water quality. High values of chemicals1 correspond to high concentrations in ammonium, azote, organic carbon, phosphorus and a high Biological Oxygen Demand, while high values of chemicals2 correspond to high concentrations in nitrate and nitrite and high conductivity, pH and suspended matter (Figure III-S4).

River topography. River distances from the outlet and from the source for each site, as well as river distance between each pair of sites, were computed using QuantumGIS software (QGIS; Quantum GIS Development Team 2017). Elevation for each site was obtained from the French Theoretical Hydrological Network ('Réseau Hydrologique Théorique français'; (Pella *et al.* 2012). A PCA was performed on elevation and distance from the outlet using R package "ade4" (Dray & Dufour 2007). The coordinates of each site on the first axis, accounting for 92.99% of the variance, were used to create a synthetic variable, hereafter named *isolation*, with high values corresponding to sites of high altitude, located far from the outlet. Additionally, the cumulative altitude differences between each pair of sites along the riverine network were computed using MATLAB software-coding environment (Mathworks, Inc., scripts available upon request). River width, used as a proxy for habitat area and hence carrying capacity (Raeymaekers *et al.* 2008), was characterised by measuring river bed width at two randomly selected locations for each sampling site, and subsequently computing the mean of these two values. The betweenness centrality value of each site was computed using ComplexNetGIS toolbox in ArcGIS (Caschili 2010). Betweenness centrality is an index quantifying the connectivity and positional importance of a node within a network (Freeman 1977; Estrada & Bodin 2008).

Table III-1: General predictions (and underlying processes) regarding the influence of environmental variables on intraspecidific α-diversity (a) and on intraspecidific β-diversity.

(a) Environmental variables	**Expected influence on intraspecific α-diversity**
Habitat heterogeneity	Highly heterogeneous sites should harbour populations with higher phenotypic α-diversity. Neutral genetic α-diversity should not be affected.
Chemicals1 Chemicals2 Oxygen saturation Temperature	Stressful conditions (i.e. high concentrations of chemicals, low oxygen saturation, high temperature) should reduce phenotypic α-diversity by strengthening selective pressures. They should also reduce effective population size and hence both neutral genetic and phenotypic α-diversity through genetic drift.
Habitat area	Populations living in large habitats should harbour high population sizes and experience low genetic drift (Prunier et al. 2017), hence increasing both neutral genetic and phenotypic α-diversity.
Connectivity	Sites with high connectivity should receive a high proportion of migrants and hence harbour populations with higher genetic and phenotypic α-diversity.
Isolation	Highly isolated sites should suffer higher genetic drift relatively to gene flow (Fourtune et al. 2016), hence reducing both neutral genetic and phenotypic α-diversity.
(b) Environmental variables	**Expected influence on intraspecific β-diversity**
Habitat dissimilarity Difference in chemicals1 Difference in chemicals2 Difference in oxygen saturation	Sites with highly dissimilar abiotic conditions should display high phenotypic β-diversity due to divergent selection. Additionally, if gene flow between environmentally different sites is hindered by the maladaptation of immigrants (isolation-by-environment, Sexton et al. 2014), we also expect a high genetic β-diversity between environmentally dissimilar sites.
Difference in habitat area	Heterogeneity in the intensity of genetic drift between sites due to contrasting population sizes should increase both neutral genetic and phenotypic β-diversity (Prunier et al. 2017).
Difference in connectivity	Dissimilarities in the intensity of gene flow experienced by populations due to contrasting connectivities should increase both neutral genetic and phenotypic β-diversity (Prunier et al. 2017).
Riverine distance Cumulative altitude difference	Sites highly isolated one from each other should experience a decrease of the homogeneizing effect of gene flow and an increase of genetic drift between them (isolation-by-distance, Hutchinson and Templeton 1999), hence enhancing both genetic and phenotypic β-diversity.

Statistical analyses

Means of intraspecific genetic and phenotypic α- and β-diversities were compared between species using Wilcoxon rank sum test. Spearman rank correlations and Mantel tests were used to assess and statistically test the significance of the correlations between genetic

and phenotypic diversity, at the α- and β-levels respectively.

The d-sep test (Shipley 2000, 2013) was used to unravel the relationships between environmental characteristics, topographical variables and intraspecific phenotypic and genetic diversity at the α- and β-levels. The d-sep test is a type of path (causal) analysis method computing the significance and likelihood of a causal model through the test of the conditional independences (named d-separations; Pearl & Verma 1987) that should be true if the model fits the data. A non-significant p-value associated with the null hypothesis "the model fits the data" indicates that the observed data are consistent with the tested model. This method is very flexible, as the statistical method used to test the independences is selected according to the data. Prior to analyses, environmental variables were log-transformed if needed to obtain a normal distribution, and all variables were centred to the mean and scaled.

At the α-level, we defined a causal model in which intraspecific genetic and phenotypic diversity were both linked one to the other and linked to oxygen saturation, water temperature, habitat heterogeneity, *chemicals1*, *chemicals2*, connectivity, *isolation*, and habitat area. As some of the topological and environmental variables are expected to covary spatially, paths taking into account these covariations were included when needed. This model was tested using a d-sep test in which d-separations (i.e. path coefficients) were tested using linear regressions. As this model did not fit the data well (see Results), it was simplified by removing paths one by one until reaching the model with the lowest Akaike Information Criteria (a measure of parsimony, Burnham & Anderson 2002) so as to identify the main determinants underlying phenotypic and genetic α-diversity.

For genetic and phenotypic β-diversity, four environmental variables (oxygen saturation, temperature, *chemicals1* and *chemicals2*) were converted into pairwise environmental differences between sites using euclidean distances. Differences in habitat area, used as a proxy for carrying capacity and hence for the effect of genetic drift, were computed as the distances based on the inverse between sites (di; Relethford 1991) as recommended in (Prunier *et al.* 2017). Additionally, we considered the three variables already in the pairwise matrix format: topographic distances, differences in cumulative altitude and habitat dissimilarities between sites. We defined a model in which genetic and phenotypic diversity were both linked one to the other and linked to these eight explanatory variables. A full model was tested using a d-sep test in which d-separations were tested using Multiple Regression on distance Matrices (MRM; Lichstein 2007), a linear regression method based on permutations. This model was simplified using the same procedure as above, until reaching the model with

the lowest AIC score.

As a side objective aiming at better understanding the spatial distribution of phenotypic diversity in the two fish species, we investigated phenotype-environment relationships by assessing and testing relationships between the individual shape of fish and raw environmental variables. For the sake of clarity, only the first two relative warps (encompassing 31.3% and 15.5% of the variance in *G. occitaniae* and 27% and 21.9% in *P. phoxinus*) were separately considered in this analysis combining model selection and model averaging. Global models (one per relative warp and per species) linking relative warps to the environmental variables and their associated quadratic terms were implemented using the *lme* function in R package 'nlme' (Pinheiro *et al.* 2016) with the population identity included as a random-intercept effect. We also added individual centroid size and its quadratic term as explanatory variables to take the effects of allometry into account (Outomuro & Johansson in press). All possible models were generated from the global model and their AIC were computed using the *dredge* function in the R package 'MuMIn' (Bartoń 2016). Full model averaging was then applied across the best models ($\Delta AIC < 4$; Burnham & Anderson 2002) with the function *model.avg* in order to estimate parameter importance and estimates. All statistical analyses were performed with the R software (R Development Core Team 2017).

III.5 - Results

Alpha- and β-intraspecific diversity

Both genetic and phenotypic α-diversity were higher for *P. phoxinus* than for *G. occitaniae* (Wilcoxon rank sum tests, $W = 149$, $P < 0.001$ for genetic α-diversity; $W = 340$, $P < 0.001$ for phenotypic α-diversity; Figures III-3a and 3b), indicating that minnow populations were on average more diverse genetically and phenotypically than gudgeon populations. However, within-species, we did not find significant correlations between genetic and phenotypic α-diversity (i.e. α-GPIDCs) for any of the two species (Spearman rank correlation tests, $\rho = 0.105$, $P = 0.521$ in *G. occitaniae* and $\rho = 0.016$, $P = 0.927$ in *P. Phoxinus*; Figures III-4a and 4b).

Mean between-sites genetic β-diversity was in average lower in *G. occitaniae* than in *P. phoxinus* (Wilcoxon rank sum test, $W = 245$, $P < 0.001$; Figure III-3c), whereas the reverse held true for mean phenotypic β-diversity per site (Wilcoxon rank sum test, $W = 1284$, $P <$

0.001; Figure III-3d); gudgeon populations were -in average- less genetically differentiated than minnow populations but more phenotypically differentiated. The fact that *G. occitaniae* populations were more phenotypically differentiated was confirmed using F_{β} (Wilcoxon rank sum test, W = 1259, P < 0.001; Figure III-S5); we further found that phenotypic turnover (measured as F_{pturn}) was also higher for *G. occitaniae* populations (Wilcoxon rank sum test, W = 981, P < 0.001; Figure III-S5). Remarkably, for 116 out of 741 population pairs of gudgeon, F_{β} and F_{pturn} were equal 1, indicating no overlap in the portions of the shape space occupied by populations, whereas in *P. phoxinus*, no pair of populations had values of F_{β} and F_{pturn} equal to 1. The correlation between genetic and phenotypic β-diversity was positive and significant in *G. occitaniae* (i.e. significant β-GPIDC, Mantel test, r = 0.358, P = 0.001; Figure III-4c) but not in *P. phoxinus* (Mantel test, r = -0.011, P = 0.521; Figure III-4d).

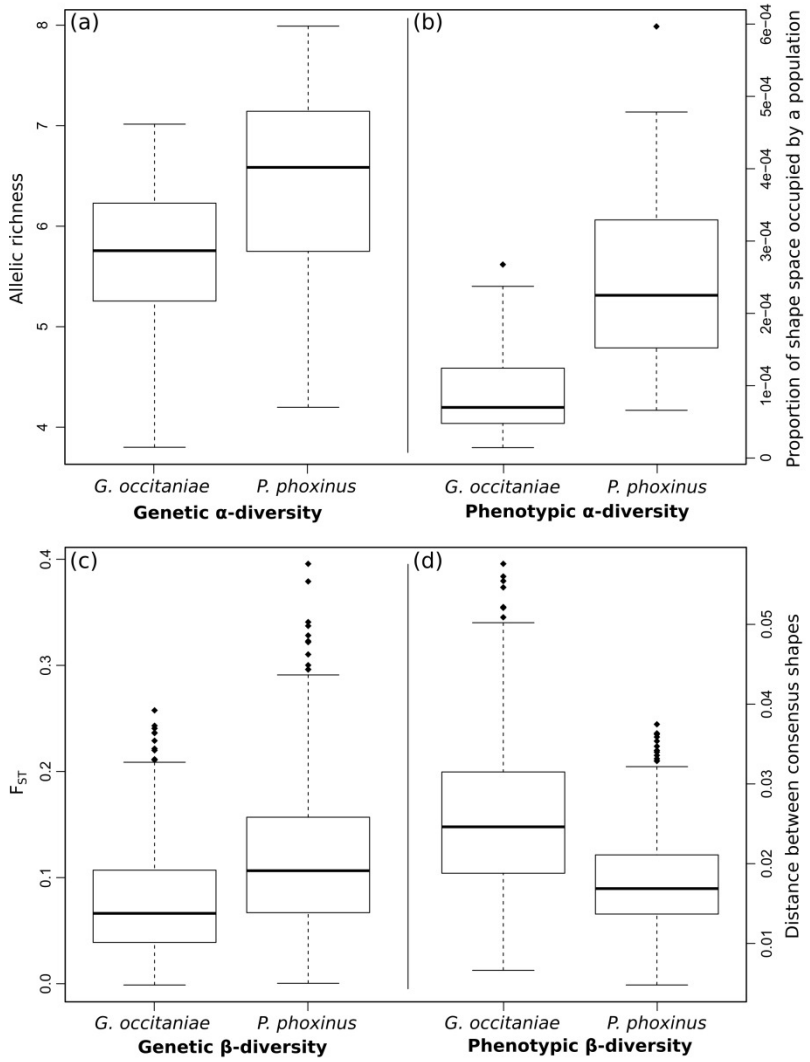


Figure III-3: Boxplots summarizing the genetic α-diversity (allelic richness) (a), phenotypic α-diversity (proportion of shape space occupied by each population) (b), genetic β-diversity (F_{ST}) (c) and phenotypic β-diversity (euclidean distance between the consensus shapes of each pair of populations) (d) in *Gobio occitaniae* and *Phoxinus phoxinus*. The solid line within each box marks the median; the length of the box is the interquartile range (from the first to the third quartile). The lower whisker extends to the first quartile minus 1.5 times the interquartile range; the upper whisker extends to the third quartile plus 1.5 times the interquartile range. Diamonds represent the data points which are beyond the whiskers.

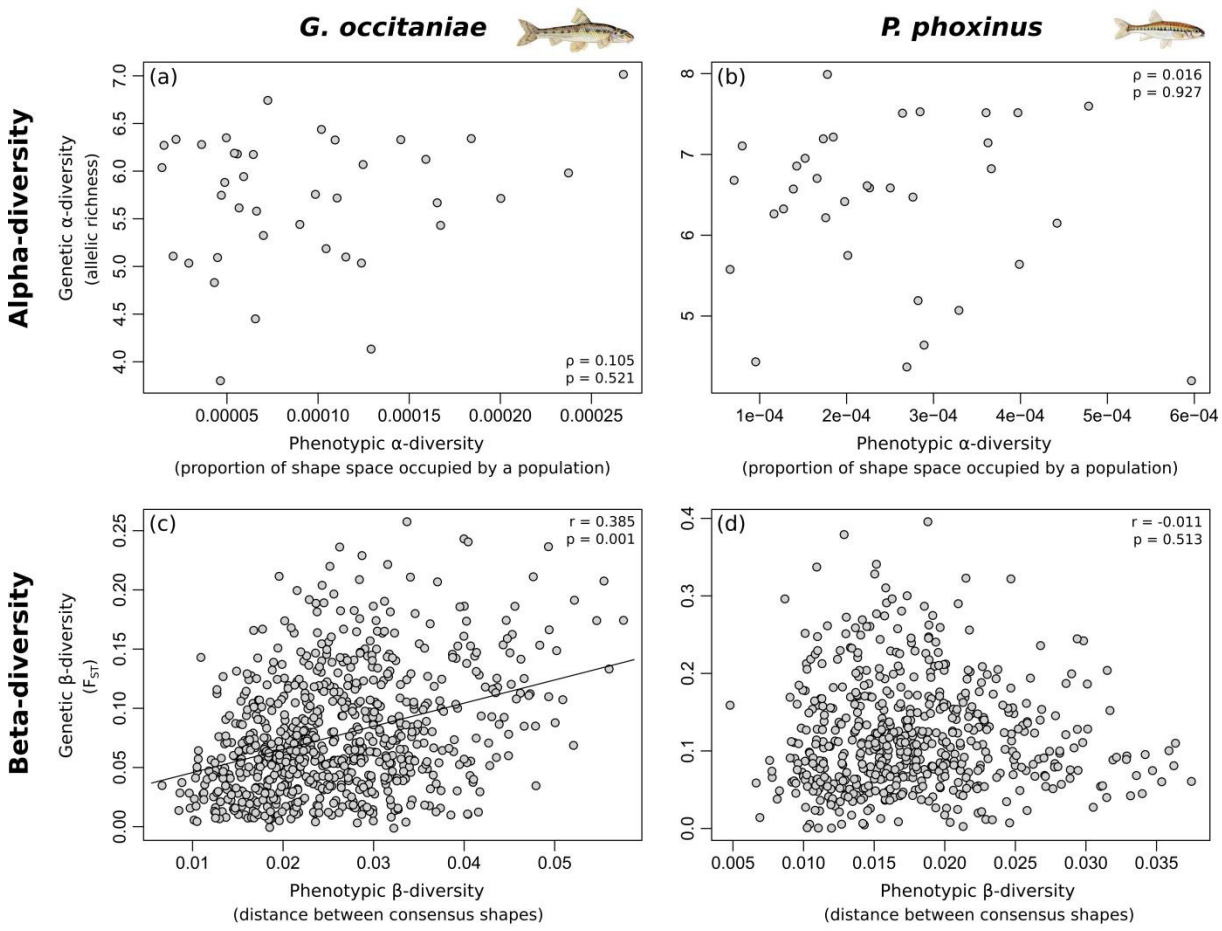


Figure III-4: Genetic α-diversity (allelic richness) of Gobio occitaniae (a) and Phoxinus phoxinus (b) plotted against phenotypic α-diversity (proportion of shape space occupied by each population) with Spearman's rho and associated P-values; and genetic β-diversity (FST) of Gobio occitaniae (c) and Phoxinus phoxinus (d) plotted against β-diversity (euclidean distance between the consensus shapes of each pair of populations) with Mantel's r and associated P-values.

Determinants of α- and β-GPIDCs.

α-GPIDCs. In *G. occitaniae* and *P. phoxinus*, the models with the lowest AIC scores were well supported by the data, as indicated by non-significant p-values (C = 75.485, 74 d.f., P = 0.389 for *G. occitaniae* and C = 76.524, 72 d.f., P = 0.335 for *P. phoxinus*, Table III-2a). In *G. occitaniae*, we found a negative effect of isolation on both genetic and phenotypic α-diversity, indicating that populations were genetically and phenotypically impoverished in sites situated at high altitude and far from the river mouth. Additionally, phenotypic α-diversity tended to be negatively related to connectivity (Figure III-5a). In *P. phoxinus*, genetic α-diversity was also negatively related to isolation, but not phenotypic diversity (Figure III-5b). Genetic α-diversity was also positively related to habitat area. Phenotypic α-

diversity was negatively correlated to oxygen saturation, which was in turn positively related to isolation and habitat area (Figure III-5b).

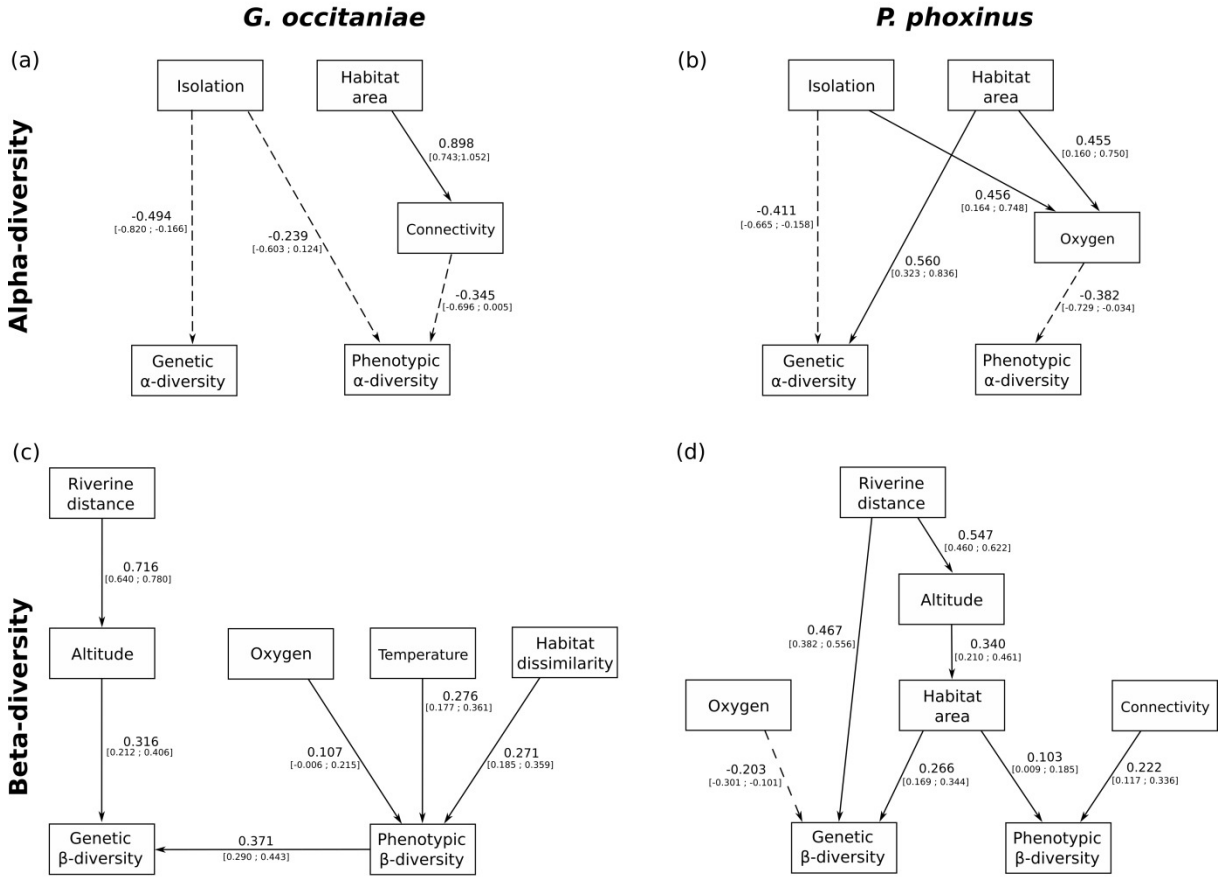


Figure III-5: Graphical representations of the models describing the causal relationships between environmental variables and genetic and phenotypic α-diversity in *Gobio occitaniae* (a) and *Phoxinus phoxinus* (b), and between environmental variables and genetic and phenotypic β-diversity in *Gobio occitaniae* (a) and *Phoxinus phoxinus* (b), obtained using the d-sep test. Single-headed arrows indicate a causal path. Solid and dashed lines stand for positive and negative values, respectively.

β-GPIDCs. The models with the lowest AIC scores were well supported by the data in both species (C = 109.167, 92 d.f., P = 0.130 for *G. occitaniae* and C = 97.699, 94 d.f., P = 0.575 in P. phoxinus, Table III-2b). For *G. occitaniae*, genetic β-diversity was positively related to the cumulative difference in altitude, which was itself related to riverine distance (leading to an indirect relationship between genetic β-diversity and riverine distance, Figure III-5c). Phenotypic β-diversity was positively correlated to three environmental variables (difference in oxygen concentration, difference in water temperature and habitat dissimilarity, Figure III-5d). Additionally, we found a positive relationship between genetic and phenotypic β-diversity (Figure III-5d). Regarding *P. phoxinus*, genetic β-diversity was positively related

to riverine distance both directly and indirectly through difference in altitude and difference in habitat area. Genetic β-diversity was also negatively related to difference in oxygen. Phenotypic β-diversity was directly related to difference in connectivity and indirectly related to pairwise riverine distance through difference in altitude and habitat area (Figure III-5d).

Table III-2: D-sep test statistics used to disentangle the effects of environmental variables on genetic and phenotypic α-diversity (a) and on genetic and phenotypic β-diversity (b) in *Gobio occitaniae* and *Phoxinus phoxinus*. For each species and diversity facet, we simplified a full model (i.e. a model including all paths described in the main text) until reaching the models with the lowest AIC score represented in Figure III-5.

(a) Alpha-intraspecific diversity	C statistics	d.f.	p-value	AIC
Gobio occitaniae				
Complete model	32.569	30	0.342	128.569
Optimal model	75.485	74	0.430	116.791
Phoxinus phoxinus				
Complete model	51.971	30	0.008	147.971
Optimal model	76.524	72	0.335	122.524
(b) Bêta-intraspecific diversity	C statistics	d.f.	p-value	AIC
Gobio occitaniae				
Complete model	47.866	30	0.020	167.866
Optimal model	108.119	92	0.120	150.119
Phoxinus phoxinus				
Complete model	50.105	30	0.012	170.105
Optimal model	97.058	94	0.394	133.058

Phenotype-environment relationships.

In *G. occitaniae*, the first relative warp had high values in individuals living in sites with high concentration in oxygen (β = 0.272, CI = [0.138; 0.406]) and low water temperature (β = -0.236, CI = [-0.367; -0.105]), and where the proportion of silt in the substrate was low (β = -0.268, CI = [-0.401; -0.136, Table III-3). Additionally, the first relative warp was related to individual centroid size (used as a proxy for individual size) and its quadratic term (β = 0.513, CI = [0.469; 0.557] and β = -0.106, CI = [-0.136; -0.076] respectively), suggesting allometry. We found no explanatory variables linked to the second relative warp among the environmental variables considered. In *P. phoxinus*, the first and second relative warps were significantly impacted by individual centroid size (β = 0.455, CI = [0.392; 0.518] and β = -

0.342, CI = [-0.413; -0.271] respectively, Table III-3), here again indicating allometry. Quite surprisingly, we found no phenotype-environment relationships in *P. phoxinus*, suggesting that, in this species, most of the phenotype variations were independent of environmental characteristics and topography.

Table III-3: Coefficients estimates and significance obtained through full model averaging on the best (ΔAIC < 4) linear mixed-effects models (one per relative warp and per species) linking relative warps to the environmental variables and their associated quadratic terms models.(*** : p-value < 0.001)

Environmental variables	Relative warp 1	Relative warp 2	Relative warp 1	Relative warp 2
	Gobio occitaniae		*Phoxinus phoxinus*	
Centroid size	0.513***		0.455***	-0.342***
Centroid size²	-0.106***			
Connectivity			-0.017	
Distance from the mouth				0.018
Distance from the source			-0.025	
Habitat area			-0.016	
Slope		0.009		0.134
Carbone	0.008	-0.048		-0.002
Carbone²		-0.063		0.007
Conductivity				-0.035
Biological Oxygen Demand			0.007	
Suspended matter				0.013
pH		0.010		
pH²		0.009		
Oxygen saturation	0.272***	0.089	0.001	0.005
Oxygen saturation²			0.006	
Temperature	-0.236***			
Silt	-0.268***	-0.016		
Cobble	-0.023			
Large cobble		0.021		
Boulder		0.007		
Large boulder		0.045		0.080

III.6 - Discussion

In this study, we tested for Genetic-Phenotypic Intraspecific Diversity correlations (GPIDCs) in two sympatric freshwater fish species, and explored the processes shaping the

spatial distribution of their genetic and phenotypic facets of diversity. Our results revealed disparities in the distribution of genetic and phenotypic diversity in the two species, as well as common and contrasted processes shaping diversity at the α- and β-level.

At the genetic level, we found that *P. phoxinus* populations were more locally diverse (higher neutral genetic α-diversity) and more differentiated (higher neutral genetic β-diversity) than *G. occitaniae* populations. The overall higher genetic α-diversity in *P. phoxinus* may indicate different evolutionary histories between the two species. For instance, this may indicate that ancient effective population sizes were higher and more stable over time in *P. phoxinus* than in *G. occitaniae*, and/or that multiple glacial refugia existed in *P. phoxinus*, hence favouring the maintenance of a high neutral genetic diversity (Hewitt 1999). Additionally, the higher genetic diversity found in *P. phoxinus* populations may indicate a lower impact of genetic drift in this species, for instance because a higher level of biological connectivity in this species than in *G. occitaniae* (Frankham 1996). However, this hypothesis is less likely given that at the β-level, populations were more genetically differentiated in *P. phoxinus* than in *G. occitaniae*, suggesting higher gene flow in *G. occitaniae* than in *P. phoxinus*. This result was not particularly expected given that populations with higher effective population sizes are expected to exchange more migrants (Dias 1996) although the lower body size of minnow may limit its dispersal capacities.

Despite these differences, we found common processes driving genetic diversity in both species. First, we found that -as expected- neutral genetic α-diversity was strongly related to geographic isolation in both species, with lower genetic diversity observed in highly isolated sites, i.e. sites at high altitude and far from the river mouth. This decrease in neutral genetic diversity in geographically isolated sites has already been reported, and has actually been suggested to be a general pattern in riverine networks (Paz-Vinas *et al.* 2015). Two non-exclusive hypotheses can explain this pattern. First, movements between populations might be directionally-biased due to water flow (Morrissey & de Kerckhove 2009; Paz-Vinas *et al.* 2013). This asymmetric dispersal leads to an increase in gene flow from upstream (isolated sites) to downstream that leads to a loss of genetic diversity upstream through emigration (Kawecki & Holt 2002). Alternatively, a decrease of genetic diversity in upstream sites might reflect the species colonization history from downstream glacial refugia. We further found that neutral genetic α-diversity in *P. phoxinus* was lower in sites of small habitat area, indicating a loss of genetic α-diversity due to small population sizes and an increased effect of genetic drift in habitat with small carrying capacity. This relationship between neutral genetic

α-diversity and habitat area was not observed in *G. occitaniae*, probably because higher gene flow in this species (lower genetic differentiation) allows counteracting the influence of drift in smaller populations.

Second, genetic β-diversity was driven by topographic features in both species; in *G. occitaniae*, genetic differentiation was higher between sites isolated from each other by high altitude drops along the network, whereas in *P. phoxinus*, genetic differentiation was higher between sites separated by a high riverine distance. These two later patterns suggest a process of isolation-by-distance (Hutchison & Templeton 1999) in the two species. Additionally, in *P. phoxinus*, we observed a positive impact of differences in habitat areas on genetic β-diversity, indicating that -as expected- populations experiencing contrasted intensity of genetic drift were genetically differentiated (Prunier *et al.* 2017), which reinforces the conclusion that this species may be more affected by genetic drift than *G. occitaniae.*

At the phenotypic level, our findings suggest that the regional pool in *G. occitaniae* was composed of poorly diverse local populations (low phenotypic α-diversity) that were highly dissimilar from one site to another (high phenotypic β-diversity with high turnover between populations, i.e. different populations display different phenotypes). Conversely, in *P. phoxinus*, phenotypic α-diversity was higher and phenotypic β-diversity were lower (than in *G. occitaniae*), which suggests that the regional pool of *P. phoxinus* was composed of highly diverse local populations that were highly similar from one site to another. The contrasted morphological patterns found in these two sympatric species may result (i) from higher effective population sizes in *P. phoxinus* than in *G. occitaniae* (which is what genetic data suggest, see above), and/or (ii) from stronger effects of selection (or environmental effects in general) in *G. occitaniae* than in *P. phoxinus*. Indeed, a stronger effect of selection is expected to lead to environmental filtering and hence less phenotypically diverse populations at the local scale (local adaptation) as well as to a high phenotypic β-diversity between populations resulting from adaptive divergence (and/or strong plastic effects) (Blanquart *et al.* 2013). This later hypothesis was strengthened by the significant relationships found between the individual shapes in *G. occitaniae* and three environmental variables (see below). The fact that *P. phoxinus* populations were highly similar from one site to another was surprising given the strong environmental heterogeneity measured among sites. This may indicate a high level of generalism in *P. phoxinus* populations, contrasting with the high level of specialism exhibited by *G. occitaniae* populations.

In line with this result, we found highly contrasted processes shaping phenotypic diversity in both species. In *G. occitaniae*, phenotypic α-diversity was lower in highly-connected sites, with a high centrality index. This result was surprising since highly central sites are expected to receive more dispersers, hence enhancing phenotypic diversity and impeding local adaptation. However, the observed pattern could be explained by a higher efficiency of selection in central sites in which dispersal introduces new phenotypes necessary for adaptation (Lenormand 2002), potentially in combination with a habitat matching process, that would hinder the negative impact of gene flow on local adaptation (Edelaar *et al.* 2008). Alternatively -and not-exclusively- this negative relationship could arise from a statistical bias, for instance if an un-measured collinear variable explains both centrality and phenotypic α-diversity. However, phenotypic α-diversity tended to be lower in isolated sites (in which, according to our former hypothesis, populations are expected to be less locally adapted and hence more diverse), which may suggest an effect of neutral processes ("phenotypic" drift) as observed in neutral genetic diversity, and/or stronger effects of environmental filtering in isolated sites (high altitude and far from the river mouth) than in less isolated sites. This latter hypothesis is likely given that upstream (isolated) sites are known to experience harsh environmental conditions (Vannote *et al.* 1980). Furthermore, in *G. occitaniae*, phenotypic β-diversity was primarily shaped by environmental variables related to habitat and water features (namely, difference in oxygen saturation, temperature and habitat dissimilarity) such that mean phenotype was different between sites displaying contrasted abiotic conditions. This impact of environment on phenotype was strengthened by the direct relationships found between individual phenotype and oxygen saturation, temperature and proportion of silt in the habitat. These two results confirm the hypothesis that selection (or environment in general) has strong effects on phenotype in *G. occitaniae*, however it remains unclear whether these effects originate from heritable differentiation or environmentally induced plasticity.

In *P. phoxinus*, phenotypic α-diversity was higher in sites with low oxygen concentration, suggesting a positive influence of stressful conditions on phenotypic α-diversity. This result was surprising as we expected that a low saturation in oxygen would sustain small population sizes, hence reducing phenotypic diversity. Moreover, stressful conditions were expected to strengthen selection pressure. However, stressful conditions have already been proven to have a positive effect on intraspecific diversity, notably (i) when they lead to an increase of mutation and recombination rates in non-neutral parts of the genome (Badyaev 2005), and (ii) when they lead to an increase in phenotypic plasticity (Ghalambor *et*

al. 2007; Rey *et al.* 2016). Phenotypic β-diversity was increased between populations inhabiting sites of different area and different connectivity. These relations may suggest an effect of neutral processes linked to population sizes or gene flow and affecting phenotypic diversity, which is likely as local adaptation does not appear to be high in this species.

We found no GPIDCs at the α-level in either species, indicating that neutral genetic α-diversity and phenotypic α-diversity are independent one from each other. Although consistent with our theoretical expectations, this result was surprising in *G. occitaniae* as we found a similar impact of isolation on genetic and phenotypic α-diversity. This absence of correlation suggests that the influence of other processes (related to connectivity) were strong enough to break spatial covariation between these two facets of diversity in this species.

At the β-level, we found a significant and positive GPIDC in *G. occitaniae*, such that populations being genetically different were also phenotypically different. However, this correlation did not seem to originate from similar environmental processes shaping both facets of β-diversity but appeared to be mainly caused by a direct effect of one facet of β-diversity on the other. It was not possible to statistically determine the direction of this relation (genetic diversity to phenotypic diversity or phenotypic diversity to genetic diversity) due to methodological limitations. However, given that we focused on neutral genetic markers, a direct impact of genetic β-diversity on phenotypic β-diversity seems unlikely except if we assume (i) that microsatellite markers chosen here reflect properly the genomic diversity in this species and (ii) that phenotypic diversity in this species is mainly driven by the genetic background of individuals. Alternatively, positive assortative mating (i.e. the propensity to mate with phenotypically similar individuals) has been shown to be particularly strong in fish (Jiang, Bolnick & Kirkpatrick 2013) and could explain this direct relation between phenotypic and genetic differentiation (Wang & Summers 2010). An unmeasured abiotic or biotic factor impacting both facets of β-diversity could also caused the observed relation, but this hypothesis seems unlikely as our dataset encompasses the main environmental variables known to be involved in adaptive and neutral processes in freshwater fish. In *P. phoxinus*, genetic and phenotypic β-diversity were not correlated despite of a similar impact of habitat area on both facets of diversity. Other important processes involving riverine distance and connectivity could impede spatial covariation between these two facets of diversity in this species.

The use of an integrative framework has allowed us to unveil striking dissimilarities between the patterns and drivers of genetic and phenotypic intraspecific diversity in two sympatric freshwater fish species. First, we found indications of strong local adaptation despite high gene flow in *G. occitaniae* populations. This result could indicate that gene flow, although often hypothesized to be negatively related to local adaptation (Räsänen & Hendry 2008), might here favour adaptation through an input of new phenotypes facilitating selection (Lenormand 2002). Second, we observed that, in *P. phoxinus*, populations were phenotypically more diverse and that gene flow was weak. This high phenotypic diversity could indicate a bet-hedging strategy (i.e. the augmentation of phenotypic diversity to optimize fitness in varying environments), possibly in response to inter-annual variations in local flow regimes (Lytle & Poff 2004). Studying neutral genetic diversity and phenotypic diversity within an integrative framework hence appeared as a valuable way of deciphering the complex and diverse impacts of neutral and adaptive processes on intraspecific diversity.

While introducing the novel framework of Species-Genetic Diversity Correlation, Vellend (2005) stated that treating interspecific and intraspecific diversity as independent phenomena in community ecology and population genetics was irrelevant. Similarly, genetic and phenotypic diversity are clearly interrelated but are mainly studied separately in population genetics and functional ecology. We advocate for a greater integration across disciplinary boundaries in future studies in order to enhance our understanding of the distribution of intraspecific diversity.

III.7 - References

Adams, D.C. & Otárola-Castillo, E. (2013) geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, **4**, 393–399.

Aljanabi, S.M. & Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research*, **25**, 4692–4693.

Badyaev, A.V. (2005) Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 877–886.

Bartoń, K. (2016) *MuMIn: Multi-Model Inference*.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

Blanchet, S., Prunier, J.G. & De Kort, H. (2017) Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics*, **33**, 579–580.

Blanchet, S., Rey, O., Etienne, R., Lek, S. & Loot, G. (2010) Species-specific responses to landscape fragmentation: implications for management strategies. *Evolutionary Applications*, **3**, 291–304.

Blanquart, F., Kaltz, O., Nuismer, S.L. & Gandon, S. (2013) A practical guide to measuring local adaptation. *Ecology Letters*, **16**, 1195–1205.

Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., Rudolf, V.H.W., Schreiber, S.J., Urban, M.C. & Vasseur, D.A. (2011) Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, **26**, 183–192.

Bolnick, D.I., Svanbäck, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D. & Forister, M.L. (2003) The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist*, **161**, 1–28.

Bookstein, F.L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge ; New York, NY.

Buisson, L., Blanc, L. & Grenouillet, G. (2008) Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish*, **17**, 244–257.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edition. Springer Verlag, New York, NY.

Caschili, S. (2010) ComplexNetGIS: a tool for the analysis of complex spatial networks. *Informatica e pianificazione urbana e territoriale* (eds G. Las Casas, P. Pontrandolfi & B. Murgante), pp. 233–242. Libria, Melfi.

Chave, J. (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters*, **16**, 4–16.

Crispo, E. & Chapman, L.J. (2008) Population genetic structure across dissolved oxygen regimes in an African cichlid fish. *Molecular Ecology*, **17**, 2134–2148.

Dias, P.C. (1996) Sources and sinks in population biology. *Trends in Ecology & Evolution*, **11**, 326–330.

Dray, S. & Dufour, A.-B. (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.

Edelaar, P., Siepielski, A.M. & Clobert, J. (2008) Matching Habitat Choice Causes Directed Gene Flow: A Neglected Dimension in Evolution and Ecology. *Evolution*, **62**, 2462–2472.

Estrada, E. & Bodin, Ö. (2008) Using network centrality measures to manage landscape connectivity. *Ecological Applications*, **18**, 1810–1825.

Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J.G. & Blanchet, S. (2016) Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. *Freshwater Biology*, **61**, 1830–1845.

Frankham, R. (1996) Relationship of Genetic Variation to Population Size in Wildlife. *Conservation Biology*, **10**, 1500–1508.

Freeman, L. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**, 35–41.

Fronhofer, E.A. & Altermatt, F. (2017) Classical metapopulation dynamics and eco-evolutionary feedbacks in dendritic networks. *Ecography*.

Ghalambor, C.K., McKAY, J.K., Carroll, S.P. & Reznick, D.N. (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, **21**, 394–407.

Grant, E.H.C., Lowe, W.H. & Fagan, W.F. (2007) Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, **10**, 165–175.

Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics*, 4th edition. Sinauer Associates, Sunderland, MA.

Hedrick, P.W. (2006) Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacy, R.C. & Dasmahapatra, K.K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, **111**, 3775–3780.

Holderegger, R., Kamm, U. & Gugerli, F. (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**, 797–807.

Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N. & Vellend, M. (2008) Ecological consequences of genetic diversity. *Ecology Letters*, **11**, 609–623.

Hutchison, D. & Templeton, A. (1999) Correlation of pairwise genetic and geographic distance measure: inferring the relative influences of gene flow and drift on distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Jiang, Y., Bolnick, D.I. & Kirkpatrick, M. (2013) Assortative Mating in Animals. *The American Naturalist*, **181**, E125–E138.

Jost, L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jung, V., Albert, C.H., Violle, C., Kunstler, G., Loucougaray, G. & Spiegelberger, T. (2013) Intraspecific trait variability mediates the response of subalpine grassland communities to extreme drought events. *Journal of Ecology*, **102**, 45–53.

Kawecki, T.J. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kawecki, T.J. & Holt, R.D. (2002) Evolutionary consequences of asymmetric dispersal rates. *The American Naturalist*, **160**, 333–347.

Lamy, T., Laroche, F., David, P., Massol, F. & Jarne, P. (2017) The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos*, **126**, 759–771.

Leimar, O. (2005) The Evolution of Phenotypic Polymorphism: Randomized Strategies versus Evolutionary Branching. *The American Naturalist*, **165**, 669–681.

Leinonen, T., McCairns, R.J.S., O'Hara, R.B. & Merilä, J. (2013) QST–FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, **14**, 179–190.

Lenormand, T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Levin, S.A. (ed). (2013) *Encyclopedia of Biodiversity*, Second edition. Elsevier, Academic Press, Amsterdam.

Lichstein, J.W. (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Loreau, M. (2000) Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos*, **91**, 3–17.

Lowe, W.H., Kovach, R.P. & Allendorf, F.W. (2017) Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution*, **32**, 141–152.

Lowe, W.H. & McPeek, M.A. (2014) Is dispersal neutral? *Trends in Ecology & Evolution*, **29**, 444–450.

Lytle, D.A. & Poff, N.L. (2004) Adaptation to natural flow regimes. *Trends in Ecology & Evolution*, **19**, 94–100.

Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mimura, M., Yahara, T., Faith, D.P., Vázquez-Domínguez, E., Colautti, R.I., Araki, H., Javadi, F., Núñez-Farfán, J., Mori, A.S., Zhou, S., Hollingsworth, P.M., Neaves, L.E., Fukano, Y., Smith, G.F., Sato, Y.-I., Tachida, H. & Hendry, A.P. (2017) Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, **10**, 121–139.

Moran, E.V., Hartig, F. & Bell, D.M. (2015) Intraspecific trait variation across scales: implications for understanding global change responses. *Global Change Biology*, **22**, 137–150.

Morrissey, M.B. & de Kerckhove, D.T. (2009) The Maintenance of Genetic Variation Due to Asymmetric Gene Flow in Dendritic Metapopulations. *The American Naturalist*, **174**, 875–889.

Nei, M., Tajima, F. & Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Odling-Smee, F.J., Laland, K.N. & Feldman, M.W. (2003) *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, Princeton, NJ and Oxford.

Outomuro, D. & Johansson, F. (in press) A potential pitfall in studies of biological shape: does size matter? *Journal of Animal Ecology*.

Paz-Vinas, I. & Blanchet, S. (2015) Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, **28**, 986–994.

Paz-Vinas, I., Loot, G., Stevens, V.M. & Blanchet, S. (2015) Evolutionary processes driving

spatial patterns of intra-specific genetic diversity in river ecosystems. *Molecular Ecology*, **24**, 4586–4604.

Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G. & Blanchet, S. (2013) The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Molecular Ecology*, **22**, 3279–3291.

Pearl, J. & Verma, T. (1987) The logic of representing dependencies by directed graphs. *Proceedings of the sixth National conference on Artificial intelligence - Volume 1*, AAAI'87, pp. 374–379. AAAI Press, Seattle, WA.

Pella, H., Lejot, J., Lamouroux, N. & Snelder, T. (2012) Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. *Géomorphologie : relief, processus, environnement*, **3**, 317–336.

Pinheiro, J., Bates, D., Debroy, S. & Sarkar, D. (2016) *Nlme: Linear and Nonlinear Mixed Effects Models*.

Prunier, J.G., Dubut, V., Chikhi, L. & Blanchet, S. (2017) Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. *Methods in Ecology and Evolution*, **Early view**.

Quantum GIS Development Team. (2017) *Quantum GIS Geographic Information System*. Open Source Geospatial Foundation Project.

R Development Core Team. (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raeymaekers, J.A.M., Maes, G.E., Geldof, S., Hontis, I., Nackaerts, K. & Volckaert, F.A.M. (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. *Evolutionary Applications*, **1**, 475–488.

Räsänen, K. & Hendry, A.P. (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.

Reed, D.H. & Frankham, R. (2003) Correlation between Fitness and Genetic Diversity. *Conservation Biology*, **17**, 230–237.

Relethford, J.H. (1991) Genetic Drift and Anthropometric Variation in Ireland. *Human Biology*, **63**, 155–165.

Rey, O., Danchin, E., Mirouze, M., Loot, C. & Blanchet, S. (2016) Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. *Trends in Ecology & Evolution*, **31**, 514–526.

Rohlf, F.J. (1993) Relative warp analysis and an example of its application to mosquito wings.

Contribution to Morphometrics, Monografías / Museo Nacional de Ciencias Naturales (eds L.F. Marcus, E. Bello & A. Garcia-Valdecasas), pp. 131–160. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, CSIC, Madrid.

Rohlf, F.J. & Marcus, L.F. (1993) A revolution morphometrics. *Trends in Ecology & Evolution*, **8**, 129–132.

Rohlf, F.J. & Slice, D. (1990) Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Biology*, **39**, 40–59.

Rousset, F. (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics*, **145**, 1219.

Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675.

Sexton, J.P., Hangartner, S.B. & Hoffmann, A.A. (2014) Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? *Evolution*, **68**, 1–15.

Shipley, B. (2000) A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, **7**, 206–218.

Shipley, B. (2013) The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology*, **94**, 560–564.

Siefert, A., Violle, C., Chalmandrier, L., Albert, C.H., Taudiere, A., Fajardo, A., Aarssen, L.W., Baraloto, C., Carlucci, M.B., Cianciaruso, M.V., de L. Dantas, V., de Bello, F., Duarte, L.D.S., Fonseca, C.R., Freschet, G.T., Gaucherand, S., Gross, N., Hikosaka, K., Jackson, B., Jung, V., Kamiyama, C., Katabuchi, M., Kembel, S.W., Kichenin, E., Kraft, N.J.B., Lagerström, A., Bagousse-Pinguet, Y.L., Li, Y., Mason, N., Messier, J., Nakashizuka, T., Overton, J.M., Peltzer, D.A., Pérez-Ramos, I.M., Pillar, V.D., Prentice, H.C., Richardson, S., Sasaki, T., Schamp, B.S., Schöb, C., Shipley, B., Sundqvist, M., Sykes, M.T., Vandewalle, M. & Wardle, D.A. (2015) A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology Letters*, **18**, 1406–1419.

Szpiech, Z.A., Jakobsson, M. & Rosenberg, N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.

Taberlet, P., Zimmermann, N.E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., Niklfeld, H., Coldea, G., Mirek, Z., Moilanen, A., Ahlmer, W., Marsan, P.A., Bona, E., Bovio, M., Choler, P., Cieślak, E., Colli, L., Cristea, V., Dalmas, J.-P., Frajman, B., Garraud, L.,

Gaudeul, M., Gielly, L., Gutermann, W., Jogan, N., Kagalo, A.A., Korbecka, G., Küpfer, P., Lequette, B., Letz, D.R., Manel, S., Mansion, G., Marhold, K., Martini, F., Negrini, R., Niño, F., Paun, O., Pellecchia, M., Perico, G., Piękoś-Mirkowa, H., Prosser, F., Puşcaş, M., Ronikier, M., Scheuerer, M., Schneeweiss, G.M., Schönswetter, P., Schratt-Ehrendorfer, L., Schüpfer, F., Selvaggi, A., Steinmann, K., Thiel-Egenter, C., van Loo, M., Winkler, M., Wohlgemuth, T., Wraber, T., Gugerli, F. & IntraBioDiv Consortium. (2012) Genetic diversity in widespread species is not congruent with species richness in alpine plant communities (ed M Vellend). *Ecology Letters*, **15**, 1439–1448.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–137.

Vellend, M. (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *The American Naturalist*, **166**, 199–215.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S.W. & Garant, D. (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, **23**, 2890–2901.

Villéger, S., Grenouillet, G. & Brosse, S. (2013) Decomposing functional β-diversity reveals that low functional β-diversity is driven by low functional turnover in European fish assemblages: Decomposing functional β-diversity. *Global Ecology and Biogeography*, **22**, 671–681.

Villéger, S., Mason, N.W.H. & Mouillot, D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, **89**, 2290–2301.

Villéger, S., Novack-Gottshall, P.M. & Mouillot, D. (2011) The multidimensionality of the niche reveals functional diversity changes in benthic marine biotas across geological time. *Ecology Letters*, **14**, 561–568.

Violle, C., Enquist, B.J., McGill, B.J., Jiang, L., Albert, C.H., Hulshof, C., Jung, V. & Messier, J. (2012) The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution*, **27**, 244–252.

Wang, I.J. & Bradburd, G.S. (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649–5662.

Wang, I.J. & Summers, K. (2010) Genetic structure is correlated with phenotypic divergence

rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19**, 447–458.

III.8 - Supplementary materials for Chapter III

Table III-S1: River name, coordinates, genetic, phenotypic and environmental data in each sampling site. Due to space limitations, the table is not available in the present manuscript but can be obtained on request.

Figure III-S2: Graphical representation of phenotypic α-diversity in two theoretical populations and of the phenotypic β-diversityand F_β between them.

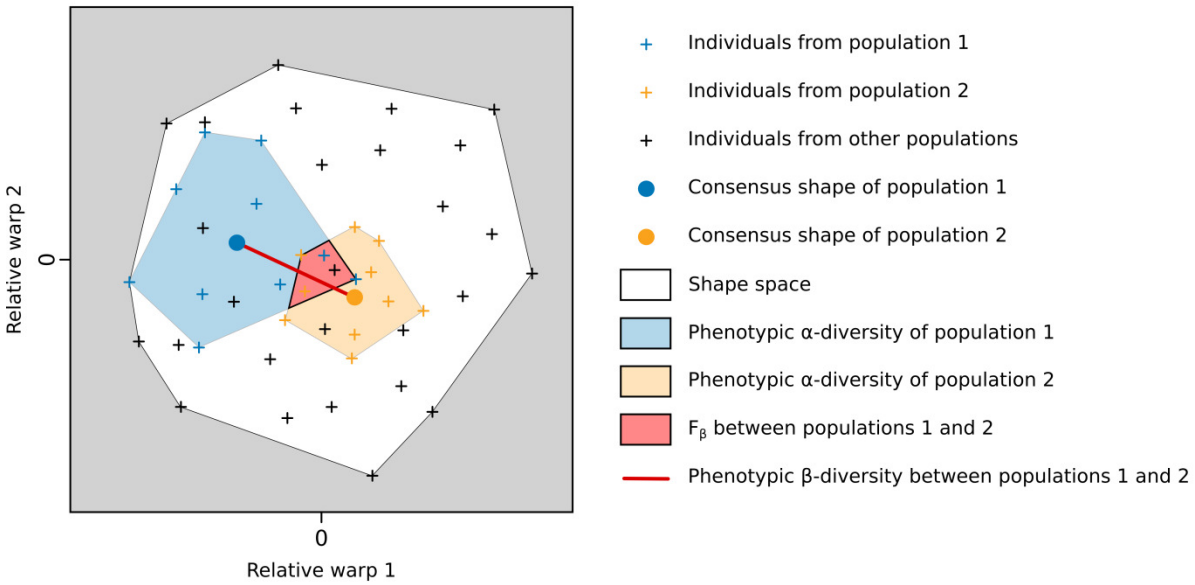


Table III-S3: Names of the nine categories of substrate and corresponding particle sizes.

Name	Particle size
Silt	< 0.2 cm
Sand	0.2-0.5 cm
Small cobble	0.5-2 cm
Cobble	2-10 cm
Large cobble	10-20 cm
Small boulder	20-40 cm
Boulder	40-60 cm
Large boulder	> 60 cm
Solid bedrock	Continuous bedrock

Figure III-S4: Correlation circle of the PCA performed on eleven chemical components.

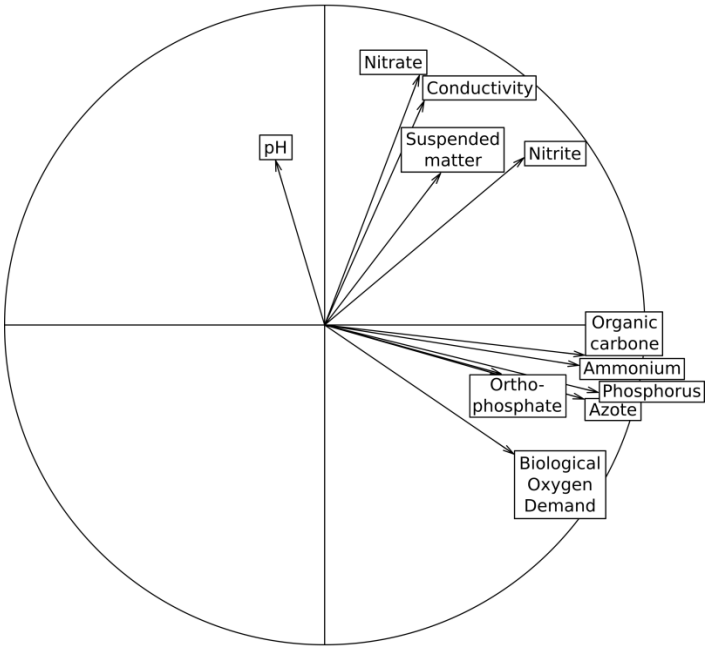
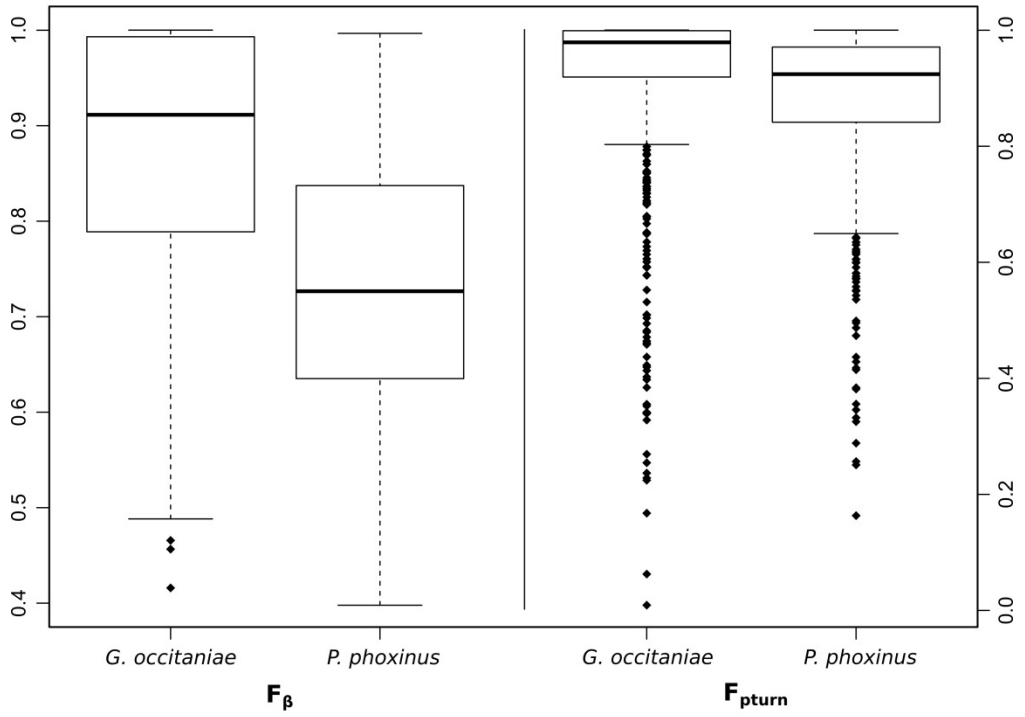


Figure III-S5: Boxplots summarizing the F_{β} and F_{pturn} in *Gobio occitaniae* and *Phoxinus phoxinus*. The solid line within each box marks the median; the length of the box is the interquartile range (from the first to the third quartile). The lower whisker extends to the first quartile minus 1.5 times the interquartile range; the upper whisker extends to the third quartile plus 1.5 times the interquartile range. Diamonds represent the data points which are beyond the whiskers.



Chapter IV - Local adaptation in riverine metapopulations: decomposing the effects of dendritic structure, environmental gradients and asymmetric dispersal.

Lisa Fourtune[1], Jérôme G. Prunier[1], Claire de Mazancourt[1], Bart Haegaeman[1], & Simon Blanchet[1,2]

[1] Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS); UMR 5321 (Station d'Écologie Théorique et Expérimentale, SETE), 09200 Moulis, France
[2] CNRS, UPS, École Nationale de Formation Agronomique (ENFA); UMR 5174 (Laboratoire Évolution & Diversité Biologique, EDB), 31062 Toulouse cedex 4, France

En préparation

IV.1 - Résumé

Au sein des métapopulations, la relation entre flux de gène et adaptation locale varie en fonction de nombreuses caractéristiques liées à la dispersion et au paysage. Etudier cette relation au sein de paysages fortement structurés comme les réseaux lotiques semble d'un grand intérêt car ces paysages sont caractérisés par une forte autocorrélation spatiale. Dans cette étude, nous avons utilisé un modèle de dynamique éco-évolutive afin d'estimer les impacts relatifs de la structure dendritique, du gradient de taille d'habitat, du gradient de conditions environnementales et de la dispersion asymétrique propres aux réseaux lotiques sur l'adaptation locale. Des simulations ont été réalisées sous différentes conditions de flux de gènes et de sélection sur migrants.Nous avons tout d'abord mis en évidence que la structure dendritique des réseaux lotiques entraine une augmentation de la connectivité des dèmes qui entraine une diminution globale de l'adaptation locale. Le gradient amont-aval de tailles d'habitat semble réduire la maladaptation dans les sites en aval, par une augmentation de la diversité génétique. Le gradient amont-aval de conditions environnementales favorise fortement l'adaptation locale, mais cet effet est réduit dans les populations périphériques via un effet de bord. La combinaison des deux gradients entraine une forte adaptation locale lorsque le flux de gènes est faible, mais qui se réduit nettement lorsque le flux de gènes augmente. Cette importance du flux de gènes provient probablement d'une interaction entre les deux gradients environnementaux qui entraine une forte augmentation de l'effet de bord quand le flux de gènes augmente. L'ajout d'une dispersion asymétrique entraine une forte augmentation de la maladaptation en aval et une diminution de l'effet de bord à l'amont du réseau. En outre, nous avons observé que les relations entre différenciation phénotypique, distance topographique et différence environnementale varient fortement selon les patrons d'adaptation locale. Nos résultats suggèrent que l'ampleur et la distribution spatiale de la maladaptation au sein des réseaux lotiques varient fortement en fonction des traits d'histoire de vie et de l'histoire évolutive des espèces considérées.

IV.2 - Abstract

Within metapopulations, the relation between gene flow and local adaptation varies depending on several dispersal-related and landscape-related features. Studying this relation within highly structured landscapes such as riverine networks appears of great interest as they are characterized by unique spatially-autocorrelated features. In this study, we used an eco-evolutionary metapopulation dynamics model in order to assess the relative impacts of dendritic structure, gradient of habitat area, gradient of environmental conditions and asymmetric dispersal rate on local adaptation within riverine networks. Simulations were run for several strengths of gene flow and selection against migrants. First, we found that the dendritic structure of riverine networks induces an increase of patches connectivity which impeded local adaptation. The increase of habitat area from upstream to downstream reduces maladaptation in downstream population through an increase of genetic diversity. The upstream-downstream gradient of environmental conditions leads to a striking increase of local adaptation which is lessened in peripheral populations through a border effect. The combination of these two gradients leads to high levels of local adaptation when gene flow is low but which strikingly decreases when the strength of gene flow increases. The importance of gene flow in this case is probably caused by an interaction between habitat areas and environmental conditions which strongly increases the magnitude of the border effect when gene flow increases. The addition of asymmetric dispersal leads to a strong increase of maladaptation in downstream sites and to a reduction of border effect upstream. Additonally, we observed that the expected relations between phenotypical differentiation, topographic distance and environmental difference strikingly vary under diverse local adaptation patterns. Our results suggest that the levels and distribution of maladaptation within riverine networks should strongly vary depending on species life-history traits and evolutionary history.

IV.3 - Introduction

Most species inhabit disparate environments and undergo spatially heterogeneous selection across their distribution (Hedrick, Ginevan & Ewing 1976; Hedrick 2006). In the absence of opposing forces, such heterogeneous selection should lead to perfect local adaptation whereby each individuals display a phenotype of optimal fitness in their local habitat. However, local adaptation is hindered by numerous evolutionary and ecological processes such as genetic drift, mutation and gene flow (Kawecki & Ebert 2004; Hartl & Clark 2007). Notably, gene flow, although indispensable for the long-term survival of meta-populations (Kokko & López-Sepulcre 2006), may introduce maladapted genes and phenotypes in populations, thus countering the diversifying selection between habitats (i.e. migration load; Slatkin 1987; Lenormand 2002). As a result, populations may be sub-optimally adapted to local conditions and the adaptive divergence between populations (i.e. the phenotypic differences that improve local fitness of populations; Räsänen & Hendry 2008) may be lowered.

The effect of gene flow on local adaptation varies depending on "dispersal-related features" (Lenormand 2002; Räsänen & Hendry 2008). First, the intensity and the spatial scale of gene flow are expected to be negatively correlated with local adaptation (i.e. strong, long-distance gene flow reduces local adaptation in all populations more than weak, short-distance gene flow; Lenormand 2002; Hanski, Mononen & Ovaskainen 2011). Second, if the survival and/or reproductive success of migrants is hindered by maladaptation (selection against migrants; Nosil *et al.* 2005), pre-adapted migrants should contribute more to gene flow than maladapted migrants, therefore reducing the pervasive influence of gene flow on local adaptation, at least partially. A similar outcome is expected if migrants tend to actively seek for and settle in habitats matching their current phenotypes (matching habitat choice; Edelaar, Siepielski & Clobert 2008; Lowe & McPeek 2014).

The impact of gene flow on local adaptation is also modulated by "landscape-related features" whereby every populations are not similarly affected by gene flow (Forester *et al.* 2016). First, the spatial arrangement of habitat patches, and the distance between habitat patches, might impede dispersal (i.e. the probability of dispersal between populations is reduced by distance; Wright 1943), thus hindering gene flow and its impact on local adaptation within isolated populations (Kawecki & Ebert 2004). Second, the heterogeneity of habitat quality or habitat area among patches can lead to a heterogeneity in population sizes

with large, high quality habitats sustaining larger populations than small, poor quality habitats (Hodgson *et al.* 2011). Populations of different sizes are not expected to be similarly impacted by gene flow. For instance, the arrival of a small amount of maladapted migrants is expected to hinder local adaptation more strongly in small populations than in large populations (Räsänen & Hendry 2008). Additionally, populations of different sizes may not contribute similarly to gene flow: all other things being equal, large populations are expected to act as sources of dispersal whereas small populations are expected to act as sinks of dispersal (Dias 1996). Consequently, small populations in low quality habitats are expected to be less locally adapted due to gene flow than populations in high quality habitats. Paradoxically, gene flow in small populations might also favour local adaptation by increasing genetic diversity or counteracting inbreeding depression (Garant, Forde & Hendry 2007). Third, the spatial distribution of environmental conditions has been theoretically proven to affect the impact of gene flow on local adaptation. For instance, Forester *et al.* (2016) showed that aggregated habitats (in which environmental conditions tend to be spatially-autocorrelated) favoured local adaptation. Additionally, the location of a population within a species range of distribution can determine how gene flow affects local adaptation. Notably, populations located at the range margins are expected to receive gene flow mainly originating from the core of the species range. The immigrants are therefore expected to be adapted to the conditions at the core range, and could strongly act against local adaptation within peripheral populations (García-Ramos & Kirkpatrick 1997; Kirkpatrick & Barton 1997; Bridle & Vines 2007).

The effective impact of gene flow on local adaptation is therefore influenced by numerous intricated factors. Several studies have aimed at empirically, experimentally or theoretically disentangle the effects of dispersal-related and/or landscape-related features on local adaptation (Räsänen & Hendry 2008). To our knowledge, the vast majority of these studies focused on terrestrial landscapes. However, studying the relation between gene flow and local adaptation within highly structured landscapes such as riverine networks appears of great interest as they are characterized by unique features (Grant, Lowe & Fagan 2007). First, riverine networks harbour a distinctive dendritic structure (Fagan 2002; Grant *et al.* 2007) consisting of a mainstem to which are typically connected an increasing number of branches from downstream to upstream. The spatial configuration of dendritic networks imposes constraints on dispersal and gene flow to freshwater species, especially in the case of species that can only disperse along the network (Grant *et al.* 2007, 2010). As a consequence, the

centrality of populations (i.e. their importance from the viewpoint of connectivity; Estrada & Bodin 2008; Eros, Schmera & Schick 2011) tends to be higher in riverine networks than within terrestrial landscapes. All other things being equal, this increase of centrality is expected to increase gene flow and its impact on local adaptation. Second, riverine networks exhibit a strong upstream-downstream gradient in habitat area, with small physical carrying capacity in upstream branches (due to small river width) and larger carrying capacity in downstream branches and mainstems (larger river width). Maladaptation is expected to be higher in small, upstream populations due to low genetic diversity that decreases the efficiency of selection (Reed & Frankham 2003). Additionally, small populations are expected to act as sinks of dispersal, further impeding local adaptation. However, dispersal in dendritic networks is also expected to be downstream-biased due to asymmetric dispersal costs caused by water flow and altitude (Morrissey & de Kerckhove 2009; Paz-Vinas *et al.* 2013). The resulting increased dispersal from upstream to downstream could counter the disequilibrium caused by population sizes and maladaptation might in that case be higher downstream than upstream (Kawecki & Holt 2002). Third, riverine networks are characterized by an upstream-downstream gradient of environmental conditions. Upstream habitats are characterized by cold, highly oxygenated water, high water velocity, a coarse grain substrate and important vegetation cover while downstream habitats are characterized by warm and poorly oxygenated water, low water velocity, a fine grain substrate and scarce vegetation cover (Vannote *et al.* 1980). Gradients of environmental conditions have been theoretically shown to hinder the negative impact of gene flow on local adaptation (Forester *et al.* 2016), which could lower the overall level of maladaptation within the network.

Studying the interactions between these features and their possible consequences on local adaptation is crucial in order to predict local adaptation patterns within riverine networks. In the present study, we used an eco-evolutionary metapopulation dynamics model (Hanski *et al.* 2011) to assess the relative impacts of the main features of riverine networks on local adaptation, for different strengths of gene flow and immigrant selection. The use of a metapopulation dynamics model was especially relevant here as metapopulation dynamics are likely to occur within riverine networks (Fronhofer & Altermatt 2017). First, we assessed the impact of the dendritic structure by comparing patterns of local adaptation within a dendritic network and a lattice network. Then, within a dendritic network, we compared the local adaptation patterns obtained when habitat area and environmental conditions were randomly-distributed or spatially-autocorrelated. Finally, in the case of spatially-autocorrelated habitat

area and environmental conditions, we tested the impact of asymmetric dispersal on local adaptation.

IV.2 - Methods

Model overview

We built on the deterministic approximation of the stochastic model of eco-evolutionary metapopulation dynamics developed by Hanski *et al.* (2011). We hereafter provide an overview of the model highlighting the features and parameters of interest in this study. For an exhaustive description of the model, see Hanski *et al.* (2011).

This continuous-time model computes the presence or absence of individuals of one species (denoted O_i for habitat patch i and taking the value of 0 or 1) in a given number of interconnected habitat patches. In each occupied patch, the model computes a mean phenotype (denoted Z_i) fluctuating through time under the influence of natural selection and gene flow. Each habitat patch is defined by an area (denoted A_i) determining its carrying capacity K_i as $prop_{AK} \times A_i$. where $prop_{AK}$ is here arbitrarily set to 0.8 (see Table IV-1 for an overview of parameters names, descriptions and default values). If a patch is occupied, it is assumed that its population size equals its carrying capacity. Each patch is also characterized by an optimal phenotype (denoted θ_i and ranging from 0.1 to 0.9 as in Hanski *et al.* 2011) that is expected to fit the environmental conditions of each habitat patch, so that an optimal phenotype maximize the fitness in a given habitat (i.e. θ_i is locally adapted to environmental conditions found in patch i). Two sites whose values of θ_i are similar are thus considered to undergo similar environmental conditions. The model assumes that environmental conditions (and hence optimal phenotypes) within each habitat patch remain constant over time (i.e. spatial heterogeneity but no temporal heterogeneity). A mismatch between the mean phenotype actually observed in patch i at a given period and the optimal phenotype of patch i (computed as $|\theta_i - Z_i|$) indicates maladaptation to environmental conditions found in patch i. Each patch is linked by branches of equal length (default length set to one) to one or more patches, and dispersal (and hence gene flow) is only possible along those branches.

The occupancy of patches varies over time according to the colonization rate of unoccupied patches C_i and the extinction rate in occupied patches E_i. When an unoccupied patch is colonized, it is assumed that all other patches can theoretically provide colonizers.

However, the contribution to colonization of each patch varies according to three main parameters. First, patches contribute to colonization according to their distance to the colonized patch (i.e. closer patches provide more migrants). The distance between patches is defined using a pairwise matrix denoted d_{ij} and the parameter α sets the range of dispersal of the species (a low value of α indicates a high probability for the modeled species to colonize distant patches and a high value of α indicates a low probability to colonize distant patches). Second, patches harbouring large population sizes (i.e. high values of A_i and K_i) contribute more to colonization than patches with small population sizes. Third, patches whose mean observed phenotypes are closer to the optimal phenotype of the colonized patch contribute further to colonization, which corresponds to a selection acting against maladapted immigrants. The strength of immigrant selection is determined by the parameter β; higher values of β correspond to a higher contribution of adapted colonizers than maladapted colonizers, i.e. stronger selection against maladapted immigrants. The newly established population then acquires the average phenotypes of the colonizers.

As stated before, once a patch has been colonized, its mean phenotype evolves according to two forces: natural selection and gene flow. Note that genetic drift is not directly taken into account by the model. Through natural selection acting at the patch level, the mean phenotype evolves towards the optimal phenotype of the patch (hence minimizing $|\theta_i - Z_i|$). The strength of local selection is defined by the parameter γ, and is also proportional to σ_G that defines the amount of genetic variance in a patch. In the original model from Hanski *et al.* (2011), σ_G was hypothesized to be equal in all patches. We here assumed that σ_G was proportional to the log value of the carrying capacity (K_i) of patches, so that selection is more efficient in patches with large population sizes (Lenormand 2002; Lanfear, Kokko & Eyre-Walker 2014). Gene flow to occupied patches is proportional to colonization rate, and is hence influenced by the same parameters (area of colonizing patches, distance from colonizing patches and maladaptation of colonizers). The parameter ρ defines the strength of gene flow, with a value of zero indicating no gene flow to occupied patches.

The extinction rate of occupied patches fluctuates over time depending mainly on two parameters. First, the extinction rate is higher in patches with small population sizes (i.e. having small values of A_i and K_i). Second, the extinction rate increases as maladaptation in patches increases, all the more so as selection at the habitat patch level is strong (i.e. for high values of γ).

After a certain amount of time, this stochastic model reaches a quasi-stationary state in

which the probability of a patch being occupied (denoted p_i* for patch i) as well as its expected mean phenotype (z_i*) become independent of time and can be easily estimated. The level of maladaptation can then be computed as $|\theta_i - z_i^*|$. Some simplifying assumptions are made in order to reach these estimations, although (Hanski *et al.* 2011) proved that the obtained estimations closely agrees with the quasi-stationary state of the stochastic model. We therefore chose to use these estimations for the sake of simplicity. A full description of the equations used here is provided in Appendix IV-S1.

Table IV-1: Name, description and default value of the model main parameters. Fixed parameters values were fixed over simulations. Parameters of interest were varied as described in the main text.

Parameter	Description	Value
Fixed parameters		
γ	Strenght of selection	5
c	Colonization rate	0.001
α	Dispersal range	2
$\alpha_{upstream}$	Dispersal range from downstream to upstream	2
K_i	Carrying capacity of patch i	$0.8 \times A_i$
σ_{Gi}	Amount of genetic variance in patch i	$0.001 \times \log(A_i)$
Parameters of interest		
ρ	Strength of gene flow	From 0 to 20
β	Strenght of immigrant selection	0, 5 and 10
d_{ij}	Distance between patch i and patch j	Varies according to configurations
A_i	Area of path i	From 10 to 100 (randomly-distributed or spatially-autocorrelated)
ϑ_i	Optimal phenotype in patch i	From 0.1 to 0.9 (randomly-distributed or spatially-autocorrelated)
$\alpha_{downstream}$	Dispersal range from upstream to downstream	From 0.2 to 2

Comparison between lattice and dendritic spatial configurations

In order to test for the impact of dendritic spatial configuration on spatial patterns of local (mal)adaptation, we compared the outputs of two spatial configurations of 36 patches each (as in Carrara *et al.* 2012; Paz-Vinas & Blanchet 2015). First, a lattice-like landscape (hereafter named "lattice configuration"), in which 36 patches are connected in a two-dimensional regular fashion (Kimura & Weiss 1964) was used to mimic a two-dimensional landscape of evenly distributed habitat patches (Figure IV-1a). Second, a dendritic-like

network (hereafter named “dendritic configuration”) was designed as 29 patches connected in a dendritic fashion continued by a seven-deme-long linear stepping-stone chain figuring the main stream of a river network (Figure IV-1b). We here assumed that the eight sites situated at the top of the network represented patches close from the sources, whereas the last patch of the chain was the closest patch from the river mouth. The upstream-downstream gradient hence comprised fourteen hierarchical levels from sources to the river mouth.

As stated above, dispersal is constrained along the branches, hence leading to two distance matrices d_{ij} describing the two spatial configurations. We modulated the distance separating two adjacent patches for each configuration separately so that mean topographic distance between patches was similar between the two configurations (i.e. the distance between two patches was 1.69 units for the lattice configuration and 1 unit for the dendritic configuration). As such, we tested the sole influence of different distance distributions and connectivity patterns stemming from different spatial configurations.

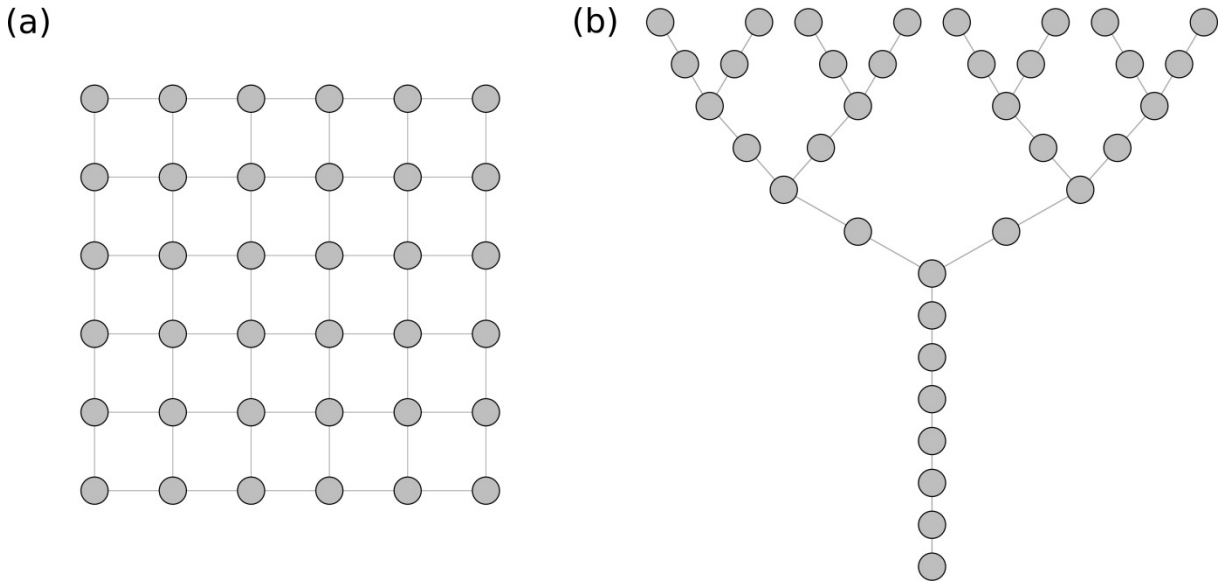


Figure IV-1: Spatial arrangement of the 36 habitat patches within the lattice configuration (a) and the dendritic configuration (b). Dispersal is constraints along the network (grey lines). Within a configuration, the lengths of all the paths are equal.

Comparison between spatially-autocorrelated and randomly-distributed patch features

To test the impact of habitat spatial autocorrelation on local adaptation, we compared outputs of simulations in which the habitat area (A_i) and/or the environmental conditions in each patch (i.e. the optimal phenotype θ_i) were either randomly-distributed or followed

upstream-downstream gradients. These gradients were composed of fourteen evenly distributed and increasing values (one per hierarchical level) ranging from 10 to 100 for A_i and from 0.1 to 0.9 for θ_i (Table IV-1).

Asymmetric dispersal range

In the previous simulations, we assumed similar dispersal ranges (parameter α) from upstream to downstream and from downstream to upstream. However, in some freshwater organisms such as fish, dispersal is expected to be downstream-biased due to asymmetric dispersal costs caused by water flow and altitude (Morrissey & de Kerckhove 2009; Paz-Vinas *et al.* 2013). Consequently, we tested the effect of asymmetric dispersal by increasing the dispersal range (by decreasing the value of α) from upstream to downstream. Two dispersal range parameters were therefore included in the model: $\alpha_{upstream}$ which was kept constant at a value of 2 and $\alpha_{downstream}$ which took values from 0.2 (high asymmetry) to 2 (no asymmetry). The effect of asymmetric dispersal was tested in the case of spatially-autocorrelated values of A_i and θ_i.

Simulation runs and summary statistics

We focused our study on the two parameters characterizing gene flow, namely ρ (the strength of gene flow) that varied from 0 (no gene flow) to 20 (high gene flow), and β (the strength of selection against immigrants) that took the following values: 0 (no immigrant selection), 5 (low immigrant selection) and 10 (high immigrant selection). When ρ is high, gene flow is expected to counteract the effect of selection, thus leading to a high level of maladaptation (high values of $|\theta_i - z_i^*|$). However, when β is high, the colonizers contributing the most to gene flow have a phenotype close to the optimal phenotype θ_i of the receiving patch, hence counteracting the negative effect of gene flow on adaptation. The other parameters were kept constant among simulations (see Table IV-1). Notably, we set the strength of local selection γ and the colonization rate c to intermediate values of 5 and 0.001 respectively (as in Hanski *et al.* 2011), and the range of dispersal α was set to 2, corresponding to a low probability for individuals from a given population to colonize patches other than those in their direct vicinity (Hanski *et al.* 2011) (Table IV-1). Reduced dispersal within riverine networks corresponds to empirical observations as well as theoretical expectations (Henriques-Silva *et al.* 2015; Fronhofer & Altermatt 2017). Patches features (habitat area A_i and habitat optimal phenotype θ_i) were randomly-distributed among patches at

each simulation run, except when they were set to be spatially-autocorrelated (see above).
For each combination of parameters, 1000 simulations were run. Over these simulations, we gathered several key results describing the system at its equilibrium and we computed their mean and variance over the 1000 simulations. First, we gathered the mean maladaptation among patches (computed as the mean of $|\theta_i - z_i^*|$ values over the 36 patches). Second, we computed the pairwise phenotypic differentiation between sites as the Euclidean distance between the z_i^* values at each site from a pair. We also gathered the difference in optimal phenotype between sites as the Euclidean distance between the values of θ_i. This difference was used as a proxy for environmental difference between patches. In order to quantify the relative influence of topography and environment on phenotypic differentiation, we computed the partial correlation coefficient between phenotypic differentiation and topographic distance (d_{ij}) while controlling for the environmental difference between patches, and the partial correlation coefficient between phenotypic differentiation and environmental difference while controlling for the topographic distance between patches, using partial Mantel tests. Third, in the simulations in which the patches followed the dendritic configuration, we aimed at describing the patterns of maladaptation within the network (i.e. determining in which patches maladaptation was highest or lowest). To that aim, we computed the correlation coefficient between maladaptation and the distance from the putative river mouth of each site using Pearson correlation test. This allowed us to determine if local adaptation followed an upstream-downstream gradient. Additionally, we computed the betweenness centrality value of each patches using the *betweenness* function from the R package 'igraph' (Csardi & Nepusz 2006) in order to quantify the centrality of each patch within the dendritic network (Freeman 1977). Highly connected patches (i.e. of high centrality) are expected to undergo higher gene flow. We computed the correlation coefficient between maladaptation and centrality using Spearman rank correlation test, in order to determine if local adaptation was promoted or hindered by connectivity. All simulations and subsequent analyses were performed using R (R Development Core Team 2017).

IV.5 - Results

Comparison between lattice and dendritic spatial configurations

As expected and over all configurations, mean maladaptation at equilibrium increased

with increasing gene flow, and decreased as selection against immigrants increased (Figure IV-2a). This confirms that mean maladaptation increases when gene flow strongly counteracts selection in each patch, thus lowering local adaptation. This result held true whatever the configuration, but maladaptation was always stronger in the dendritic configuration (light blue dots, Figure IV-2a) than in the lattice configuration (brown dots, Figure IV-2a).

The strongest correlation between phenotypic differentiation and topographic distance (after controlling for environmental difference) was observed when gene flow was high, and when selection against immigrants was null (Figure IV-2b), i.e. when phenotypic differentiation was mostly driven by gene flow rather than by natural selection. Conversely (and as expected), phenotypic differentiation was strongly driven by environmental differences (local adaptation) when gene flow was low and when selection against immigrants was high (Figure IV-2c). All other things being equal, the relationships between topographic distance and phenotypic differentiation was stronger within the dendritic configuration, whereas the relationships between environment features and phenotypic differentiation was stronger within the lattice configuration (Figures IV-2b and IV-2c).

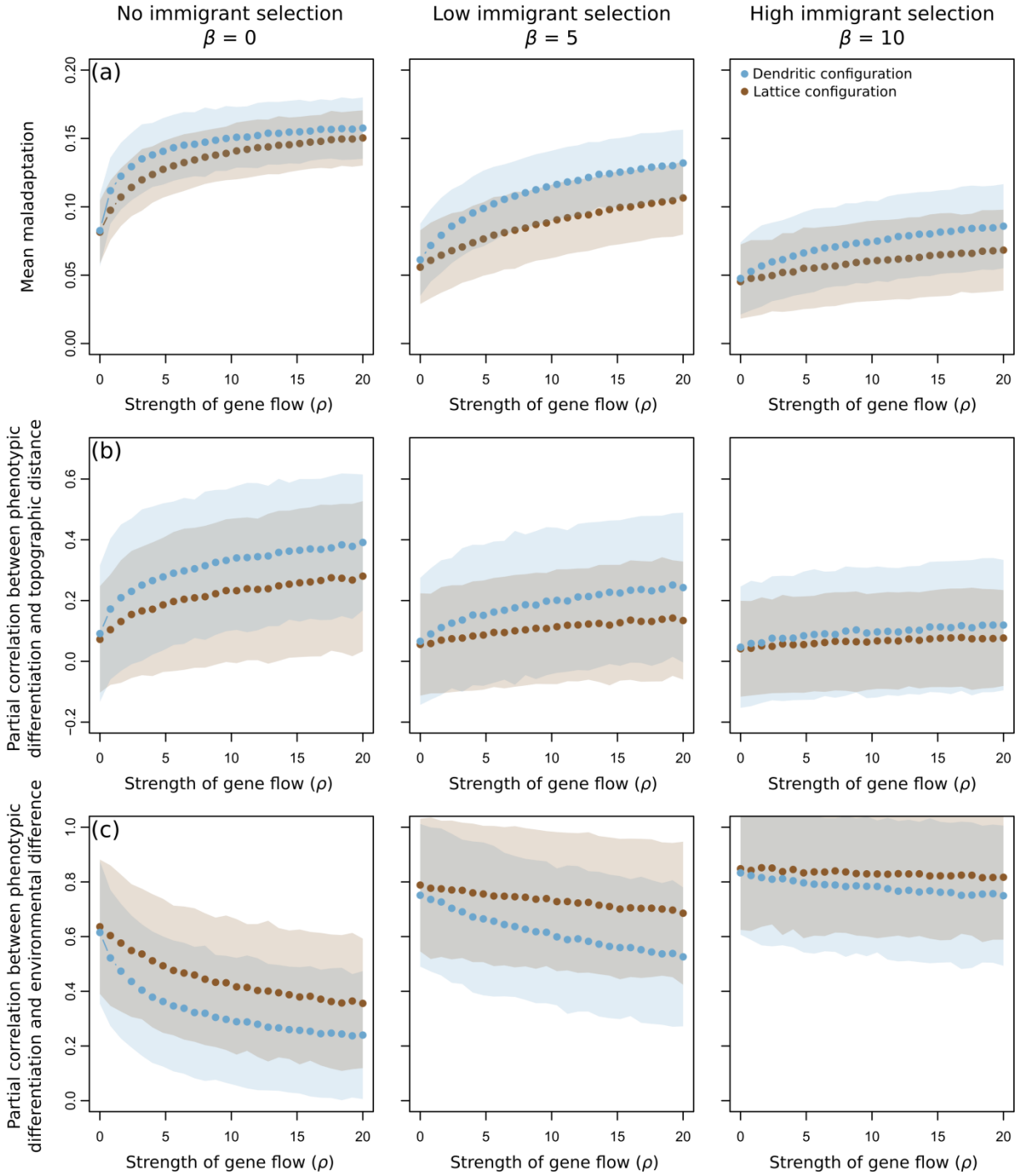


Figure IV-2: Mean maladaptation (a), partial correlation between phenotypical differentiation and topographic distance after accounting for environmental difference (b) and partial correlation between phenotypical differentiation and environmental difference after accounting for topographic distance (c) obtained within the dendritic (light blue dots) and the lattice (brown dotes) configurations, for values of strength of gene flow (ρ) ranging from 0 (no gene flow) to 20 (high gene flow), and for three values of immigrant selection (β) (0: non immigrant selection, 5: low immigrant selection, 10: high immigrant selection). Other parameters were held constant, and patch sizes and environmental conditions were randomly distributed (see Table IV-1). Dots are mean values obtained over 1000 simulations run for each combination of parameter values and confidence intervals are built from the standard deviations obtained over these 1000 simulations.

Comparison between spatially-autocorrelated and randomly-distributed habitat features

When habitat areas were spatially-autocorrelated but environmental conditions were randomly-distributed, patterns of maladaptation as well as patterns of phenotypic differentiation were very similar to what was observed when both habitat areas and environmental conditions were randomly distributed (orange dots vs. light blue dotes in Figures IV-3 and IV-4). The only notable exceptions were that correlations between maladaptation and site centrality on one side, and between maladaptation and distance from the river mouth on the other side were stronger when habitat areas were spatially-autocorrelated than when they were randomly-distributed, especially when gene flow was low (Figures IV-3b & IV-3c).

More strikingly, deviation from a non-structured landscape was stronger when environmental conditions were spatially-autocorrelated (green dots and dark blue dots in Figures IV-3 & IV-4). In particular, maladaptation was strikingly lowered when environmental conditions were spatially-autocorrelated (Figure IV-3a). Maladaptation was the weakest when habitat areas were randomly-distributed (green dots, Figure IV-3a), indirectly suggesting that the combination of spatially-autocorrelated habitat areas and spatially-autocorrelated environmental conditions may hinder local adaptation, especially when gene flow is high and when there is no selection against immigrants (dark blue dots, Figure IV-3a). Spatially-autocorrelated environmental conditions associated with random-distributed habitat areas did not strongly impacted maladaptation spatial distribution (Figures IV-3b and IV-3c). However, the combination of spatially-autocorrelated environmental conditions and spatially-autocorrelated habitat areas led to striking spatial patterns of maladaptation (Figures IV-3b & IV-3c). We observed that both centrality and distance from the mouth (both being partially correlated one to each other in our dendritic network) were strong predictors of maladaptation when both habitat areas and environmental conditions were spatially-autocorrelated, with highest levels of maladaptation being observed in isolated sites (negative correlations; Figure IV-3b), located far from the river mouth (positive correlations; Figure IV-3c). Spatially-autocorrelated environmental conditions also led phenotypic differentiation to be almost exclusively driven by environmental difference (Figures IV-4b), whereas the impact of topography on phenotypic differentiation appeared negligible (Figure IV-4a). This result held true with and without spatially-autocorrelated habitat areas.

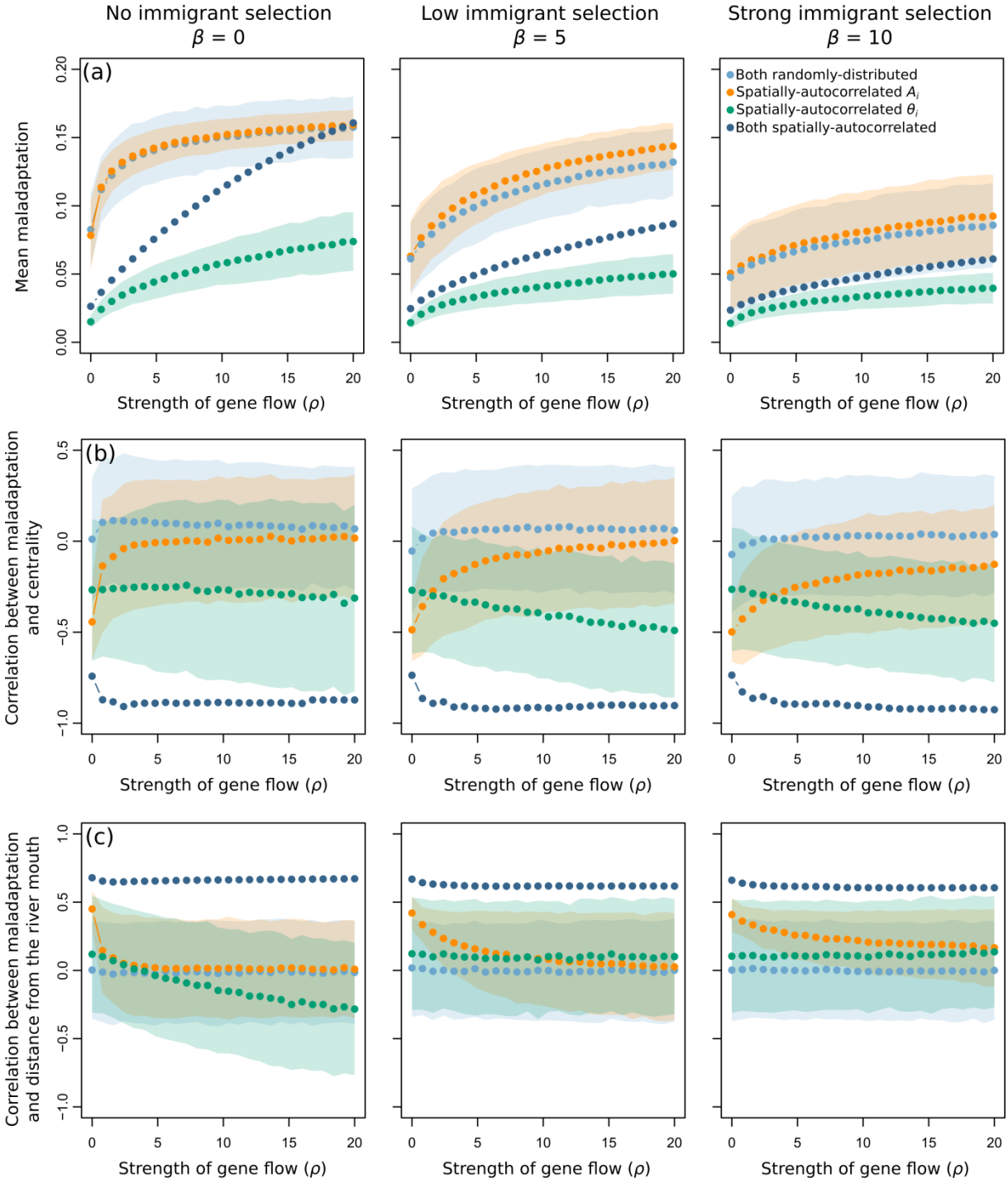


Figure IV-3: Mean maladaptation (a), correlation between maladaptation and betweenness centrality (b), and correlation between maladaptation and distance from the river mouth (c) obtained within the dendritic configuration, with randomly-distributed habitat area and optimal phenotype (light blue dots), spatially-autocorrelated habitat area and randomly-distributed optimal phenotype (orange dots), randomly-distributed habitat area and spatially-autocorrelated optimal phenotype (green dots) and spatially-autocorrelated habitat area and optimal phenotype (dark blue dots), for values of strength of gene flow (ρ) ranging from 0 (no gene flow) to 20 (high gene flow), and for three values of immigrant selection (β) (0 = non immigrant selection, 5 = low immigrant selection, 10 = high immigrant selection). Other parameters were held constant (see Table IV-1). See legends in Figure IV-2 for details.

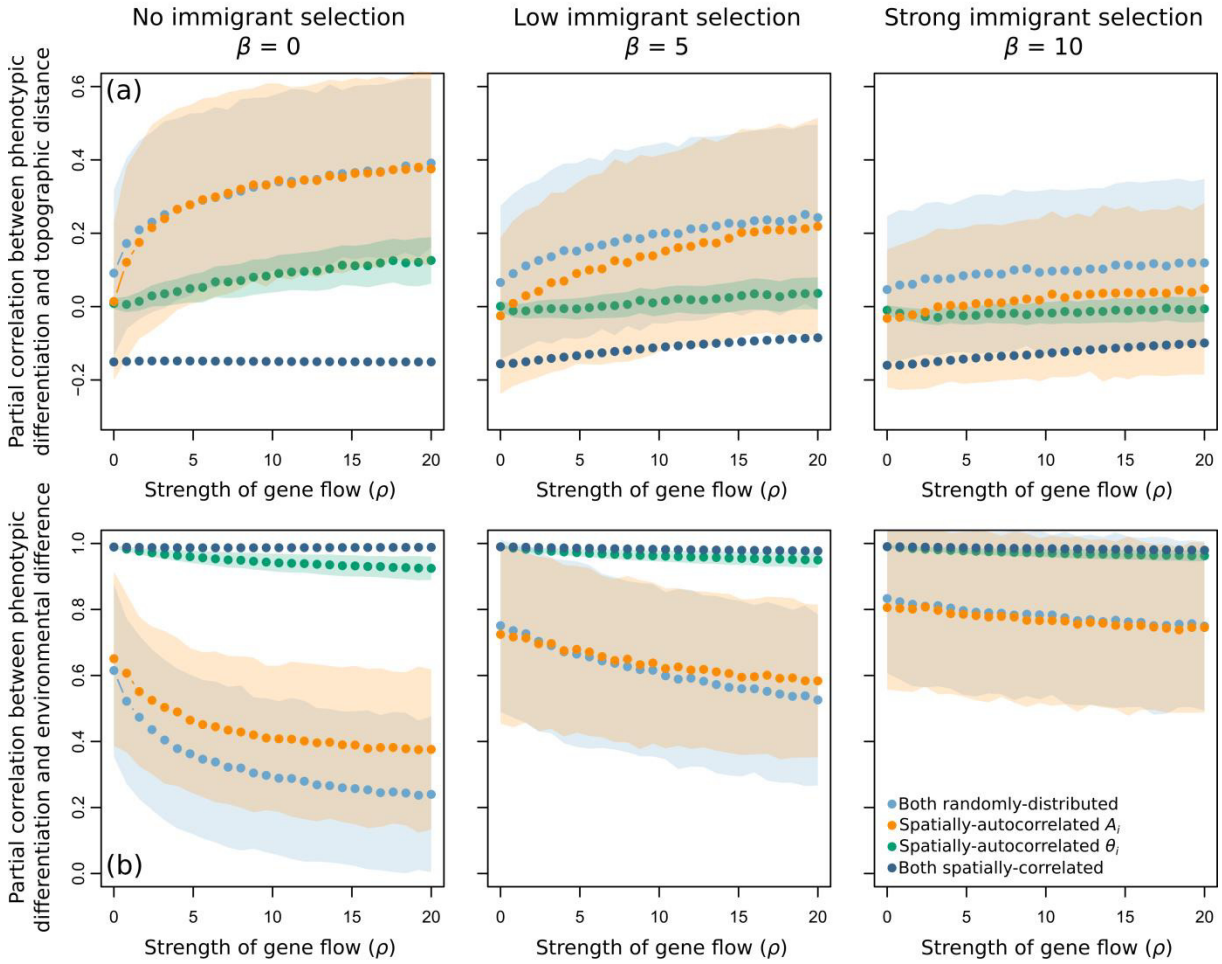


Figure IV-4: Partial correlation between phenotypical differentiation and topographic distance after accounting for environmental difference (a) and partial correlation between phenotypical differentiation and environmental difference after accounting for topographic distance (b) obtained within the dendritic configuration. See legends of Figure IV-2 and IV-3 for details.

Asymmetric dispersal range

When both habitat area and environment conditions were spatially-autocorrelated, downstream-biased gene flow (simulated through a higher dispersal range from upstream to downstream) tended to increase maladaptation (Figure IV-5a) and strikingly impacted the distribution of maladaptation within the network: contrary to what was observed with unbiased gene flow, the highest levels of maladaptation were found in downstream sites (Figure IV-5b), and well-connected sites (Figure IV-5c). Downstream-biased gene flow also appeared to increase the correlation between topographic distance and phenotypic differentiation while slightly decreasing the correlation between environmental difference and phenotypic differentiation (Figures IV-5d & IV-5e).

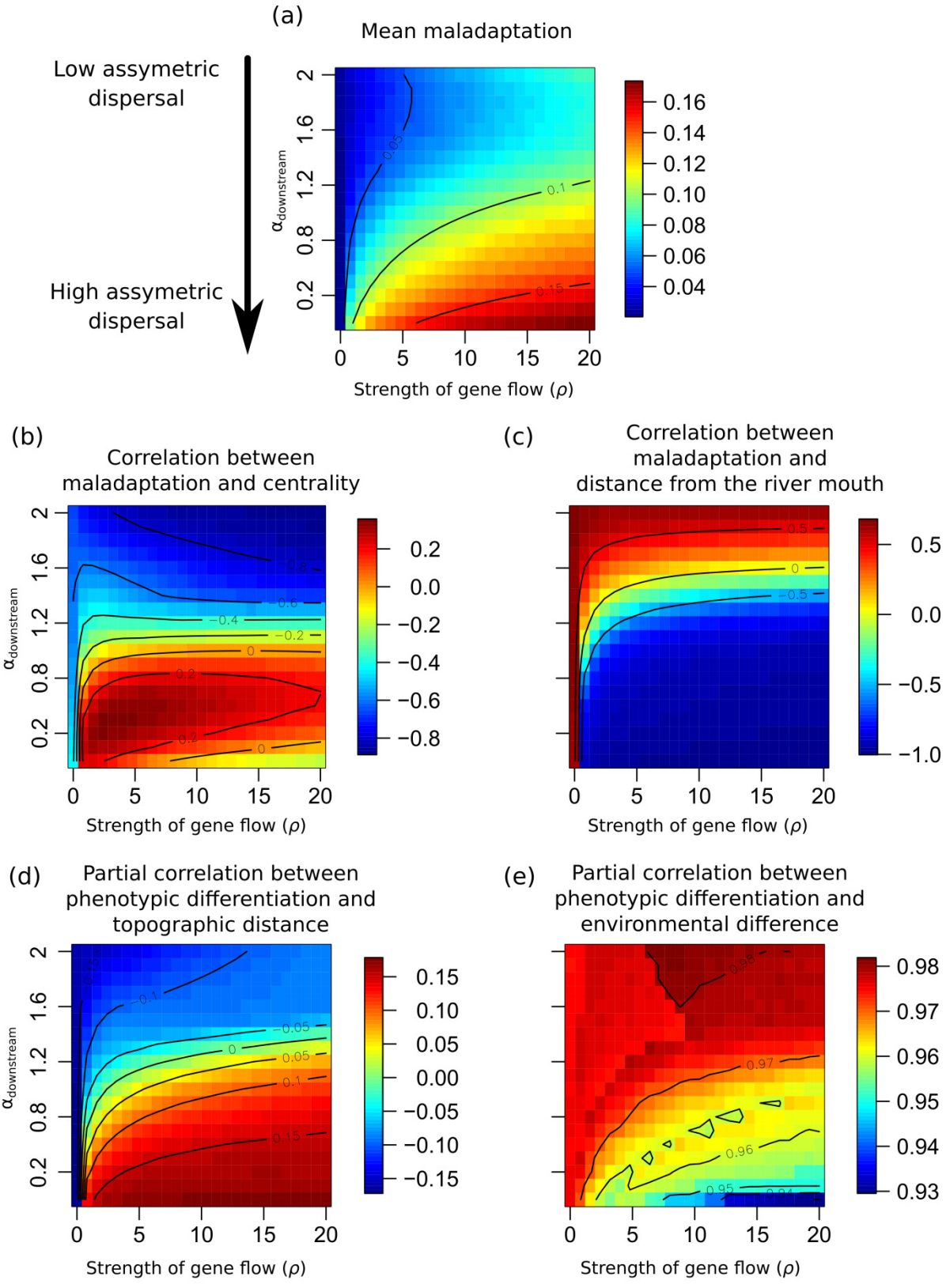


Figure IV-5: Mean maladaptation (a), correlation between maladaptation and betweenness centrality (b), correlation between maladaptation and distance from the river mouth (c), partial correlation between phenotypical differentiation and topographic distance after accounting for environmental difference (d) and partial correlation between phenotypical differentiation and environmental

difference after accounting for topographic distance (e) obtained within the dendritic configuration with spatially-distributed habitat area and optimal phenotype plotted in relation to the strength of gene flow (ρ) and the migration range from upstream to downstream ($\alpha_{downstream}$) (0.2 = high augmentation of migration range from upstream to downstream, i.e. strong asymmetric dispersal and 2 = no augmentation of migration range from upstream to downstream, i.e. no asymmetric dispersal). Other parameters were held constant (see Table IV-1) and immigrant selection (β) was set to 5. Values are means obtained over 1000 simulations run for each combination of parameter values.

IV.6 - Discussion

Our comprehension of spatial patterns of local adaptation depends on our ability to understand how selection and gene flow interact within realistic landscapes. Using an eco-evolutionary metapopulation dynamics model developed by Hanski *et al.* (2011), we were able to theoretically assess the impacts of the main features of dendritic river networks on spatial patterns of local adaptation.

First, we observed that, as expected, mean maladaptation decreased as immigrant selection increased, selection against maladapted migrants masking the effect of gene flow on local adaptation. Additionally, we found stronger mean maladaptation within the dendritic configuration than within the lattice configuration, whatever the level of immigrant selection (Figure IV-2a). This result suggests that the spatial arrangement of patches observed within dendritic networks, and the co-varying changes in patch connectivity affected the global level of maladaptation in these landscapes. This increase in maladaption in dendritic networks might be due to the increase in mean patch connectivity in these networks (mean centrality was 200 in the dendritic network and 105 in the lattice network), which tends to facilitate gene flow and hence reduce the possibility for local adaptation to occur (see Figure IV-S2).

We also observed that phenotypic differentiation was more strongly linked to environment difference in the lattice configuration, whereas it was more strongly linked to topographic distance in the dendritic configuration, reinforcing the conclusion that adaptation to local conditions is higher in lattice landscapes and that -on the contrary- connectivity is a key driver of phenotypic patterns in dendritic networks. The specific patterns of connectivity observed within riverine networks have already been shown to impact, among others, inter- and intraspecific diversity patterns (Carrara *et al.* 2012, 2014; Paz-Vinas & Blanchet 2015) as well as the evolution of dispersal within the network (Henriques-Silva *et al.* 2015; Fronhofer & Altermatt 2017), but it is, to our knowledge, the first time that its impact on local adaptation

has been highlighted.

Within the dendritic network, we observed that spatial-autocorrelation in habitat areas (i.e. the fact that habitat areas, and hence population sized increased from upstream to downstream) did not influence the mean level of maladaptation (Figure IV-3a), but modified the spatial distribution of maladaptation when gene flow was very low. In particular, maladaptation was lower downstream and in well connected sites (also mainly located downstream) (Figures IV-3b, IV-3c and IV-S3a), probably because natural selection is more efficient in large-sized populations with higher genetic diversity (up to a certain point where gene flow masked the effect of natural selection whichever the population size, which probably explained why these patterns were not observed when gene flow was increased, Figure IV-S3b) (Dias 1996; Lanfear *et al.* 2014).

When environmental conditions were spatially-autocorrelated from upstream to downstream, we observed a strong decrease in levels of maladaptation, suggesting that the gradient of environmental conditions favoured local adaptation (Figure IV-3a), which is consistent with the findings of Forester *et al.* (2016). A likely explanation for this result is that, when environmental conditions follow a spatial gradient, adjacent patches harbour individuals phenotypically similar, therefore gene flow tends to bring pre-adapted individuals. Accordingly, the levels of maladaptation observed with spatially-autocorrelated environmental conditions and with randomly-distributed environmental conditions were more similar when the selection on migrants increased (i.e. when the influence of surrounding sites was minimum). Furthermore, the correlations between centrality and maladaptation tended to be negative when environmental conditions followed an uptream-downstream gradient (see Figure IV-S4), suggesting a border effect whereby populations located at the range margins experience extreme environmental conditions (extreme values of the gradient) but receive migrants adapted to the conditions at the core range which lead to high levels of maladaptation (Bridle & Vines 2007). This specific pattern of local adaptation led to a high correlation between phenotypic differentiation and environmental conditions indicating that populations inhabiting similar environment tend to display similar phenotypes, even when they are in a permanent state of maladaptation (such as apical populations in Figure IV-S4).

When habitat areas and environmental conditions followed an upstream-downstream gradient (thus mimicking the conditions expected in nature), the level of maladaptation within the network increased strikingly with gene flow, notably when the selection against migrants was weak (Figures IV-3a and IV-S5). When gene flow is low, the gradient in environmental

conditions again appears to favour local adaptation. The border effect appears strong upstream, but is almost non-existent downstream probably because of a higher genetic diversity in downstream populations (that are large in size in this situation) that favours local adaptation (Figure IV-S5a). When gene flow increases, local adaptation is strongly hindered, which suggests that the increase in population sizes from upstream to downstream has a strong negative impact on local adaptation in these conditions. A likely explanation for this result is that, when gene flow is high, border effects are stronger because populations at the core range are of intermediate to high population sizes, and thus contribute strongly to gene flow. As above, phenotypic differentiation and environmental conditions were strongly correlated, indicating a high propensity of similar environment to harbour phenotypically similar populations.

We observed that asymmetric dispersal led to a strong increase in levels of maladaptation, especially when gene flow was high (Figures IV-5a and IV-S6). Additionally, asymmetric dispersal changed the pattern of maladaptation by leading to a strong increase of maladaptation downstream and a slight decrease in maladaptation upstream (Figures IV-5b and IV-S6b). This result suggests that asymmetric dispersal increases the masking effect of gene flow downstream thus impeding local adaptation in downstream populations to occur. The border effect observed upstream without asymmetric dispersal tends to disappear, probably because the quasi-unidirectional gene flow limits the propagation of maladapted phenotypes from the core populations to the upstream populations. Additionally, asymmetric dispersal tended to increase the influence of topography and to decrease the influence of environment on phenotypic differentiation (Figure IV-6d).

We found strong and diverse impacts of riverine networks features on local adaptation patterns. Our results suggest that the levels and distribution of maladaptation within the network should strongly vary depending on species life-history traits and evolutionary history. Notably, species having high dispersal abilities are expected to display higher levels of maladaptation, especially in peripheral populations through border effect. However, this border effect is expected to be hindered downstream if the species display higher genetic diversity downstream through high population sizes and/or post-glaciation colonization history (Paz-Vinas *et al.* 2015). Additionally, freshwater species display contrasting propensity to asymmetric dispersal: passive dispersers such as drifting macroinvertebrates are expected to display higher asymmetry in their dispersal than active dispersers such as fish (Brown *et al.* 2011). Species in which asymmetric dispersal is strong are expected to display

higher levels of maladaptation in downstream populations and lower levels of maladaptation in upstream populations than species with weak or no asymmetric dispersal. Further, we expect a strong positive correlation between phenotypic differentiation and topographic distance in species with strong asymmetric dispersal whereas this correlation is expected to be null or negative in the absence of asymmetric dispersal.

The detection of local adaptation is a major challenge in ecology and demands complex experimental studies such as transplant experiments (Blanquart *et al.* 2013). Here, we found that the expected relations between phenotypical differentiation (which can be assessed empirically), topographic distance and environmental difference strikingly vary under diverse local adaptation patterns. Gathering such correlations to indirectly estimate local adaptation strength and patterns as preliminary empirical analyses appears worth exploring.

Through the present model, we were able to disentangle the role of the main riverine network features in shaping local adaptation. However, some important processes were not considered. Notably, genetic drift is not taken into account in this model although it is expected to impede local adaptation, notably in small population (Garant *et al.* 2007). Similarly, the potential positive effect of gene flow on local adaptation (for instance through the introduction of novel alleles or phenotypes on which can act selection, Räsänen & Hendry 2008) has not been considered here. Our results therefore need to be carefully interpreted and call for further theoretical, empirical and experimental studies to be confirmed and widened.

IV.7 - References

Blanquart, F., Kaltz, O., Nuismer, S.L. & Gandon, S. (2013) A practical guide to measuring local adaptation. *Ecology Letters*, **16**, 1195–1205.

Bridle, J.R. & Vines, T.H. (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, **22**, 140–147.

Brown, B.L., Swan, C.M., Auerbach, D.A., Campbell Grant, E.H., Hitt, N.P., Maloney, K.O. & Patrick, C. (2011) Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems. *Journal of the North American Benthological Society*, **30**, 310–327.

Carrara, F., Altermatt, F., Rodriguez-Iturbe, I. & Rinaldo, A. (2012) Dendritic connectivity

controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences*, **109**, 5761–5766.

Carrara, F., Rinaldo, A., Giometto, A. & Altermatt, F. (2014) Complex Interaction of Dendritic Connectivity and Hierarchical Patch Size on Biodiversity in River-Like Landscapes. *The American Naturalist*, **183**, 13–25.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1965.

Dias, P.C. (1996) Sources and sinks in population biology. *Trends in Ecology & Evolution*, **11**, 326–330.

Edelaar, P., Siepielski, A.M. & Clobert, J. (2008) Matching Habitat Choice Causes Directed Gene Flow: A Neglected Dimension in Evolution and Ecology. *Evolution*, **62**, 2462–2472.

Eros, T., Schmera, D. & Schick, R.S. (2011) Network thinking in riverscape conservation – A graph-based approach. *Biological Conservation*, **144**, 184–192.

Estrada, E. & Bodin, Ö. (2008) Using network centrality measures to manage landscape connectivity. *Ecological Applications*, **18**, 1810–1825.

Fagan, W.F. (2002) Connectivity, Fragmentation, and Extinction Risk in Dendritic Metapopulations. *Ecology*, **83**, 3243–3249.

Forester, B.R., Jones, M.R., Joost, S., Landguth, E.L. & Lasky, J.R. (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.

Freeman, L. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**, 35–41.

Fronhofer, E.A. & Altermatt, F. (2017) Classical metapopulation dynamics and eco-evolutionary feedbacks in dendritic networks. *Ecography*.

Garant, D., Forde, S.E. & Hendry, A.P. (2007) The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, **21**, 434–443.

García-Ramos, G. & Kirkpatrick, M. (1997) Genetic Models of Adaptation and Gene Flow in Peripheral Populations. *Evolution*, **51**, 21–28.

Grant, E.H.C., Lowe, W.H. & Fagan, W.F. (2007) Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, **10**, 165–175.

Grant, E.H.C., Nichols, J.D., Lowe, W.H. & Fagan, W.F. (2010) Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. *Proceedings of the National Academy of Sciences*, **107**, 6936–6940.

Hanski, I., Mononen, T. & Ovaskainen, O. (2011) Eco-Evolutionary Metapopulation Dynamics and the Spatial Scale of Adaptation. *The American Naturalist*, **177**, 29–43.

Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics*, 4th edition. Sinauer Associates, Sunderland, MA.

Hedrick, P.W. (2006) Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.

Hedrick, P.W., Ginevan, M.E. & Ewing, E.P. (1976) Genetic Polymorphism in Heterogeneous Environments. *Annual Review of Ecology and Systematics*, **7**, 1–32.

Henriques-Silva, R., Boivin, F., Calcagno, V., Urban, M.C. & Peres-Neto, P.R. (2015) On the evolution of dispersal via heterogeneity in spatial connectivity. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**, 20142879.

Hodgson, J.A., Moilanen, A., Wintle, B.A. & Thomas, C.D. (2011) Habitat area, quality and connectivity: striking the balance for efficient conservation. *Journal of Applied Ecology*, **48**, 148–152.

Kawecki, T.J. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kawecki, T.J. & Holt, R.D. (2002) Evolutionary consequences of asymmetric dispersal rates. *The American Naturalist*, **160**, 333–347.

Kimura, M. & Weiss, G.H. (1964) The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, **49**, 561–576.

Kirkpatrick, M. & Barton, N.H. (1997) Evolution of a Species' Range. *The American Naturalist*, **150**, 1–23.

Kokko, H. & López-Sepulcre, A. (2006) From Individual Dispersal to Species Ranges: Perspectives for a Changing World. *Science*, **313**, 789–791.

Lanfear, R., Kokko, H. & Eyre-Walker, A. (2014) Population size and the rate of evolution. *Trends in Ecology & Evolution*, **29**, 33–41.

Lenormand, T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Lowe, W.H. & McPeek, M.A. (2014) Is dispersal neutral? *Trends in Ecology & Evolution*, **29**, 444–450.

Morrissey, M.B. & de Kerckhove, D.T. (2009) The Maintenance of Genetic Variation Due to Asymmetric Gene Flow in Dendritic Metapopulations. *The American Naturalist*, **174**, 875–889.

Nosil, P., Vines, T.H., Funk, D.J. & Harrison, R. (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.

Paz-Vinas, I. & Blanchet, S. (2015) Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology*, **28**, 986–994.

Paz-Vinas, I., Loot, G., Stevens, V.M. & Blanchet, S. (2015) Evolutionary processes driving spatial patterns of intra-specific genetic diversity in river ecosystems. *Molecular Ecology*, **24**, 4586–4604.

Paz-Vinas, I., Quéméré, E., Chikhi, L., Loot, G. & Blanchet, S. (2013) The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. *Molecular Ecology*, **22**, 3279–3291.

R Development Core Team. (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Räsänen, K. & Hendry, A.P. (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.

Reed, D.H. & Frankham, R. (2003) Correlation between Fitness and Genetic Diversity. *Conservation Biology*, **17**, 230–237.

Slatkin, M. (1987) Gene Flow and the Geographic Structure of Natural Populations. *Science*, **236**, 787–792.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. & Cushing, C.E. (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 130–137.

Wright, S. (1943) Isolation by Distance. *Genetics*, **28**, 114–138.

IV.8 - Supplementary materials for Chapter IV

Appendix IV-S1: Equations used in the simulations, based on the model created by Hanski, Mononen & Ovaskainen (2011)

- Growth rate in patch *i*:

$$r_i^* = 1 - \frac{\gamma}{2}(\theta_i - z_i^*)^2$$

- Time to extinction in patch *i*:

$$T_i^* = \frac{K_i^{s_i^*}}{s_i^* r_i^*}\left[1 - \frac{1 + s_i^* \ln(K_i)}{\exp\left(s_i^* \ln(K_i)\right)}\right]$$

where

$$s_i^* = 2r_i^*$$

- Extinction rate in patch *i*:

$$E_i^* = \frac{1}{T_i^*}$$

- Contribution of patch *j* to the colonization rate of patch *i*:

$$m_{ij}^* = c\, K_j\, e^{r_j^*} e^{-\beta|\vartheta_i - z_i^*|} A_i \frac{\alpha^2}{2\pi}\, e^{-\alpha d_{ij}}$$

- Probability of patch *i* of being occupied:

$$p_i^* = \frac{\sum_{j \neq i} m_{ij}^* p_j^*}{E_i^* + \sum_{j \neq i} m_{ij}^* p_j^*}$$

- Expected mean phenotype in patch *i*:

$$z_i^* = \frac{\gamma(\sigma_G \log(K_i))\theta_i + \left[E_i^* + (\rho/A_i)\sum_{j\neq i} m_{ij}^* p_j^*\right] z_i^{C*}}{\gamma(\sigma_G \log(K_i)) + E_i^* + (\rho/A_i)\sum_{j\neq i} m_{ij}^* p_j^*}$$

where

$$z_i^{C*} = \frac{\sum_{j\neq i} m_{ij}^* z_j^* p_j^*}{\sum_{j\neq i} m_{ij}^* p_j^*}$$

Figure IV-S2: Graphical representation of the 36 habitat patches within the lattice (a) and the dendritic (b) configurations, colored after their mean maladaptation computed over 10000 simulation runs (white: maladaptation of 0.102, black: maldaptation of 0.137). The strength of gene flow (ρ) was set to 20 and immigrant selection (θ) was set to 5. Habitat areas (A_i) and optimal phenotypes (ϑ_i) were randomly-distributed at each simulation run. For other parameter values see Table IV-1.

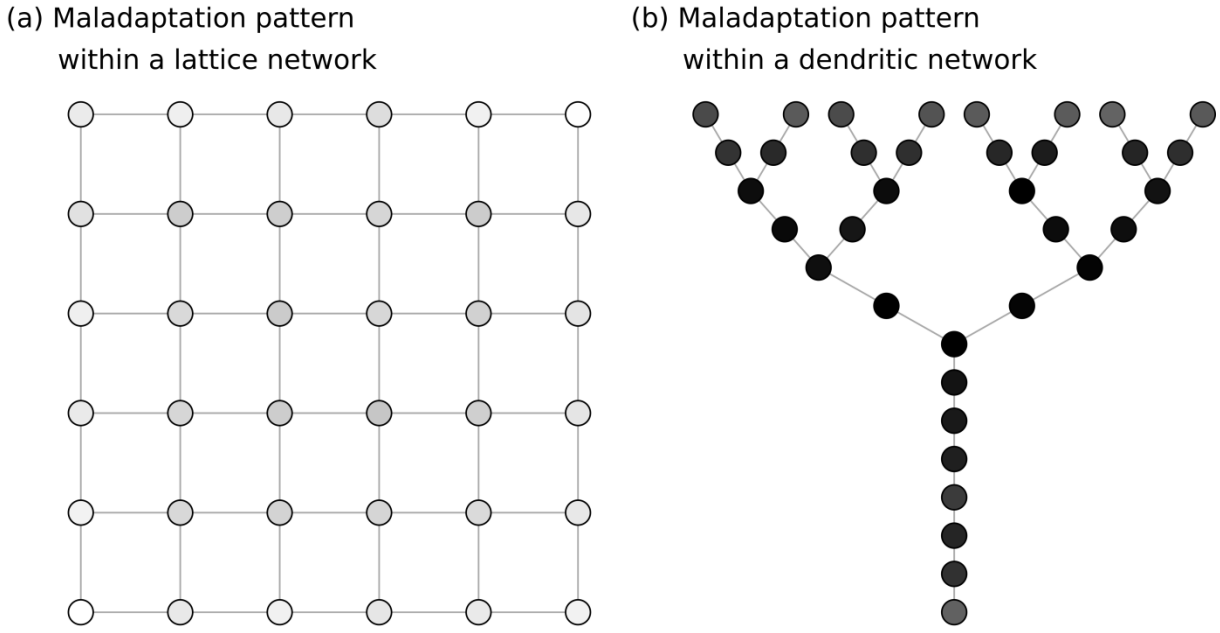


Figure IV-S3: Graphical representation of the 36 habitat patches within the dendritic configuration colored after their mean maladaptation computed over 10000 simulation runs ranging from 0.006 (white) to 0.165 (black). The strength of gene flow (ρ) was set to 0.1 (a) and 20 (b). Immigrant selection (β) was set to 5. Habitat areas (A_i) were spatially-autocorrelated and optimal phenotypes (ϑ_i) were randomly-distributed at each simulation run. For other parameter values see Table IV-1.

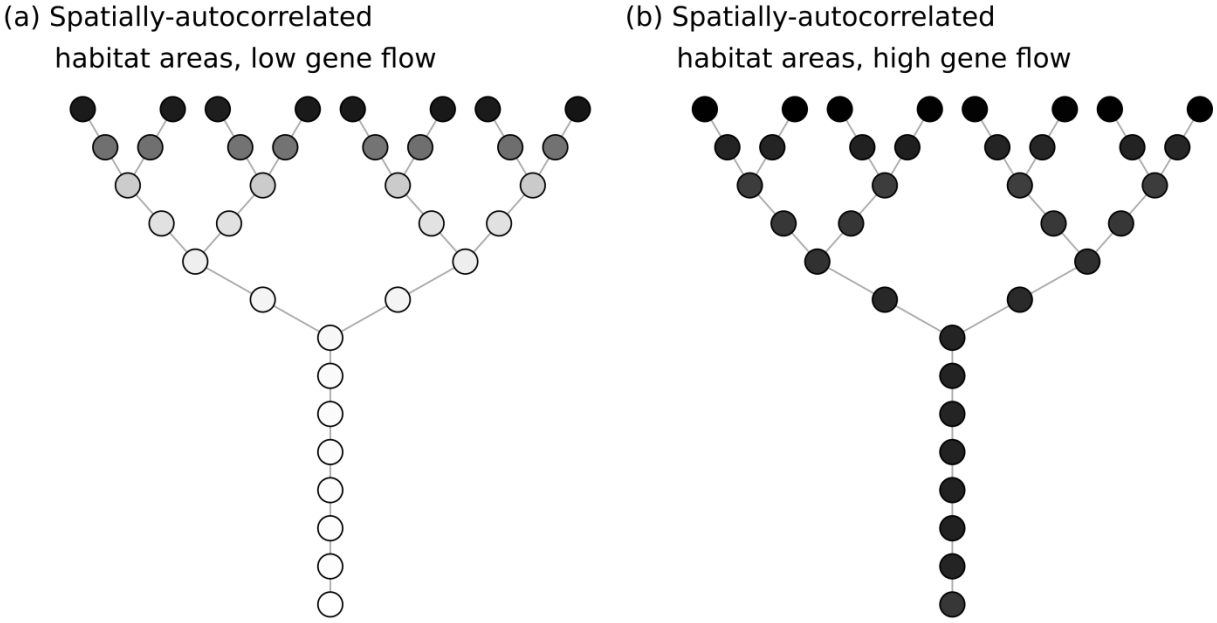


Figure IV-S4: Graphical representation of the 36 habitat patches within the dendritic configuration colored after their mean maladaptation computed over 10000 simulation runs ranging from 0.009 (white) to 0.103 (black). The strength of gene flow (ρ) was set to 0.1 (a) and 20 (b). Immigrant selection (β) was set to 5. Optimal phenotypes (ϑ_i) were spatially-autocorrelated and habitat areas (A_i) were randomly-distributed at each simulation run. For other parameter values see Table IV-1.

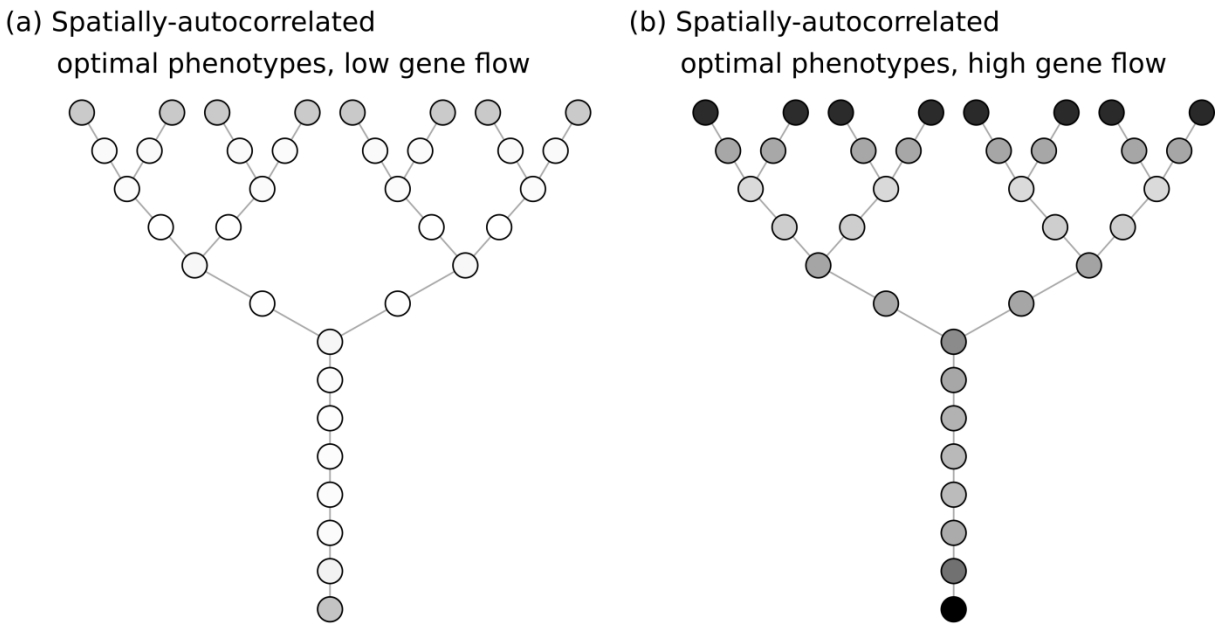


Figure IV-S5: Graphical representation of the 36 habitat patches within the dendritic configuration colored after their mean maladaptation computed over 10000 simulation runs ranging from 0.0001 (white) to 0.161 (black). The strength of gene flow (ρ) was set to 0.1 (a) and 20 (b). Immigrant selection (β) was set to 5. Habitat areas (A_i) and optimal phenotypes (ϑ_i) were spatially-autocorrelated. For other parameter values see Table IV-1.

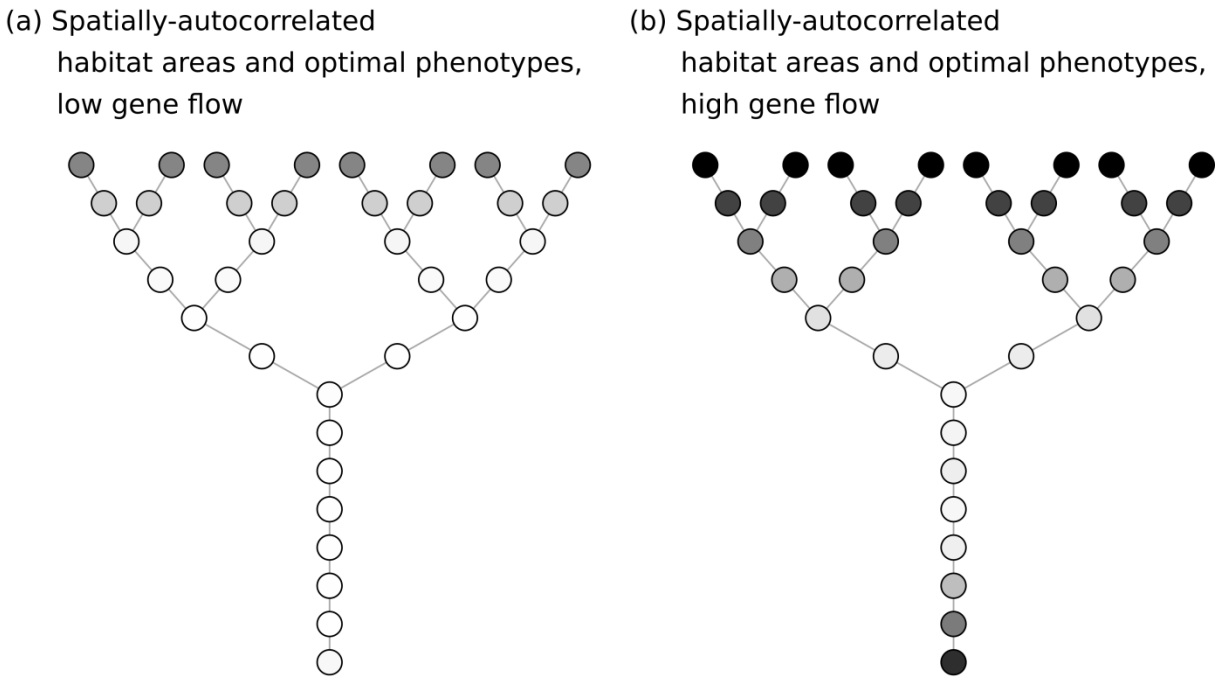
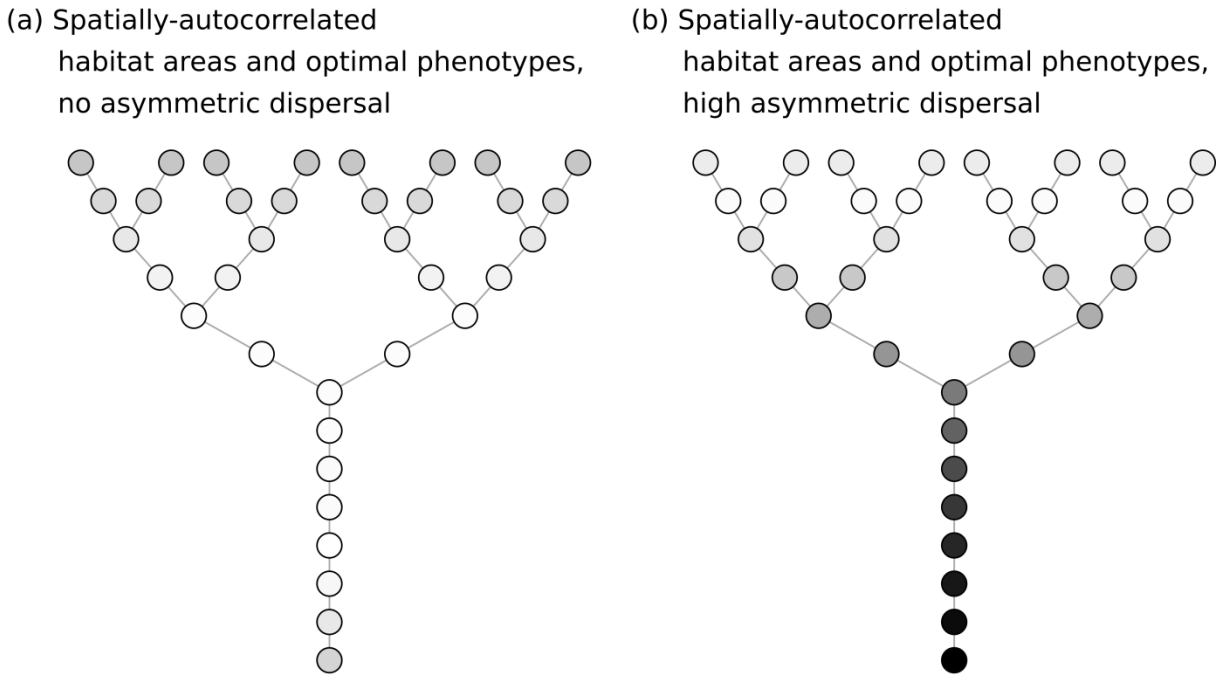


Figure IV-S6: Graphical representation of the 36 habitat patches within the dendritic configuration colored after their mean maladaptation computed over 10000 simulation runs ranging from 0.001 (white) to 0.583 (black). The migration range from upstream to downstream ($\alpha_{downstream}$) was set to 2 (no asymmetric dispersal) (a) and 0.2 (strong asymmetric dispersal) (b). The strength of gene flow (ρ) was set to 10 and immigrant selection (β) was set to 5. Habitat areas (A_i) and optimal phenotypes (ϑ_i) were spatially-autocorrelated. For other parameter values see Table IV-1.



General conclusion

Throughout the present work, my collaborators and myself aimed at uncovering how interspecific diversity, intraspecific genetic diversity and intraspecific phenotypic diversity were spatially distributed and whether they were driven by analogous or contrasted processes within riverine networks. First, we developed novel statistical approaches aiming at deciphering complex causal links among several variables in the form of distance matrices. Second, we uncovered similar patterns of species and genetic diversity of four sympatric freshwater fish species, and we showed that similar ecological and evolutionary processes related to environmental filtering, migration, drift and colonization history act for explaining both species and genetic diversity of fish communities. Third, we unveiled strikingly dissimilar patterns of genetic and phenotypic intraspecific diversity in two sympatric freshwater fish species, and we deciphered the complex and diverse impacts of neutral and adaptive processes on intraspecific diversity in these two species. Fourth, we theoretically assessed the impact of the habitat features characterizing riverine networks on local adaptation and its relation with gene flow. We notably found unexpected interactions between the gradients in carrying capacities and environmental conditions that strongly influence patterns of maladaptation within the network. The present conclusion aims at summarizing the main findings of this work.

The patterns and environmental drivers of biodiversity within riverine networks

Interspecific diversity

In Chapter II, we uncovered two main drivers of freshwater fish interspecific α-diversity. First, we found a direct and positive effect of habitat area on community diversity, which corresponds to the species-area relationship (Connor and McCoy 1979) whereby the species diversity within an ecosystem tends to increase with the area of the ecosystem. Within riverine networks, habitat area -often assessed through stream order or stream width- is a variable commonly considered when studying interspecific diversity distribution. A recent meta-analysis found that the increase of interspecific diversity in freshwater fish with habitat area was a widely observed pattern (Vander Vorste et al. 2017). Several hypotheses have been formulated to explain this relationship in riverine network. Notably, a higher habitat area

could sustain higher density through enhanced foraging and enhanced refuge from predation (Jackson et al. 2001). The mechanistic processes leading to this relation could however vary across taxa or geographic regions (Vander Vorste et al. 2017). Second, we found that interspecific diversity was higher in sites of low altitude, located close from the river mouth. As the species-area relationship, the altitudinal gradient in interspecific diversity is a widespread pattern in ecology (Rahbek 1995). In freshwater fish communities, this pattern has been observed across several geographical regions (e.g. Heino 2002; Fu et al. 2004). Downstream sites are expected to experience higher immigration through asymmetric dispersal in riverine network (Ronce and Kirkpatrick 2001), which might provide new species to these sites. Alternatively, downstream sites are expected to be environmentally more stable than sites located in high altitude, which undergo stronger interannual disturbances in discharge, turbidity and temperature (Vannote et al. 1980). Environmental stability is expected to be positively related to interspecific diversity (Whittaker et al. 2001).

At the β-level, we found that interspecific diversity was higher between sites displaying different water characteristics (e.g. water temperature and pH). When species show habitat preferences, the structure and composition of communities is expected to vary with environmental features (Jackson et al. 2001). Such environmental filtering may therefore strongly drive community differentiation, which has been demonstrated in diverse freshwater organisms, including fish (Blanchet et al. 2014).

Intraspecific diversity

Genetic intraspecific diversity. In Chapters I, II and III, we uncovered several drivers of neutral genetic α-diversity within riverine network. First, we observed that genetic α-diversity of all the species studied was lower in sites of high altitude and located far from the river mouth. At it has already been stated in previous chapters, this decrease in genetic α-diversity in geographically isolated sites has been suggested to be a general pattern in riverine networks (Paz-Vinas et al. 2015). Assymetric dispersal could explain this pattern, as it is expected to lead to an increase in gene flow from upstream (isolated sites) to downstream that leads to a loss of genetic α-diversity upstream through emigration (Kawecki and Holt 2002; Morrissey and de Kerckhove 2009). Alternatively, a decrease of genetic α-diversity in upstream sites might reflect the species colonization history from downstream glacial refugia (Paz-Vinas et al. 2015). Additionally, we observed an increase of genetic α-diversity in sites of large area in some, but not all, species. The more likely explanation for this pattern is that

sites of large areas can sustain higher population sizes, which are known to harbour higher genetic α-diversity (Frankham 1996).

At the β-level, we found in all studied species an increase of the neutral genetic differentiation between populations isolated from one another (by topographic distance and/or cumulative difference in altitude between them). This result corresponds to a pattern of isolation-by-distance whereby the homogenizing effect of gene flow decreases and the relative effect of genetic drift increases as the geographic distance between sites increases (Hutchison and Templeton 1999). This pattern was the strongest pattern of genetic β-diversity observed and was found consistently in all species. Other drivers such as the number of weirs, the difference in habitat area and the difference of water quality between populations were found to impact genetic differentiation in some, but not all, studies.

Phenotypic intraspecific diversity. Phenotypic diversity was assessed in two species (namely *Gobio occitaniae* and *Phoxinus phoxinus,* see Chapter III) and the patterns and drivers of phenotypic α- and β- diversity strikingly varies between species. We found that, in *G. occitaniae*, phenotype α-diversity was lower in highly-connected sites. This pattern could be explained by a higher efficiency of selection in central sites in which dispersal introduces new phenotypes necessary for adaptation (Lenormand 2002). Phenotypic α-diversity also tended to be lower in isolated sites, which may suggest a stronger effect of environmental filtering in these sites, given that they are known to experience harsh environmental conditions (Vannote et al. 1980). In *G. occitaniae*, phenotypic β-diversity was primarily shaped by environmental variables related to habitat and water features such that mean phenotype was different between sites displaying contrasted abiotic conditions. This result suggests that selection (or environment in general) has strong effects on phenotype in *G. occitaniae*, however it remains unclear whether these effects originate from heritable differentiation or environmentally induced plasticity.

In *P. phoxinus*, phenotypic α-diversity was higher in sites with low oxygen concentration, suggesting a positive influence of stressful conditions on phenotypic α-diversity. This result could indicate that stressful conditions have a positive effect on intraspecific diversity, notably (i) when they lead to an increase of mutation and recombination rates in non-neutral parts of the genome (Badyaev 2005), and (ii) when they lead to an increase in phenotypic plasticity (Ghalambor et al. 2007; Rey et al. 2016). Phenotypic β-diversity was increased between populations inhabiting sites of different area

and different connectivity. These relations suggest an effect of neutral processes linked to population sizes or gene flow and affecting phenotypic diversity (Prunier et al. 2017), which is likely as local adaptation does not appear to be high in this species.

Additionally, through the use of an eco-evolutionary metapopulation dynamics model we uncovered how the landscape features of riverine networks influenced local adaptation. Notably we found that the expected relations between phenotypic β-diversity, topographic distance and environmental difference strikingly vary under diverse local adaptation patterns. We observed that phenotypic β-diversity was strongly related to environmental difference when local adaptation was strong (which theoretically confirms the hypothesis of a strong local adaptation in *G. occitaniae*) and was strongly related to topographic distance when local adaptation was dampened, for example through asymmetric dispersal.

Correlation between biodiversity facets

As biodiversity facets are hypothesized to be driven by similar processes (Antonovics 1976), and to directly influence one another (Vellend and Geber 2005), we chose to search for correlations between them.

In Chapter II, we uncovered significant positive correlations between interspecific diversity and genetic intraspecific diversity (Species-Genetic Diversity Correlations, SGDC; Vellend 2005). At the α-level, the correlation between interspecific diversity and genetic diversity appeared to be caused by parallel effects of environmental conditions on both facets of diversity. Indeed, as we saw above, interspecific α-diversity and genetic α-diversity were both negatively impacted by isolation and positively impacted by habitat area.
At the β-level, we found that the positive correlation between interspecific diversity and intraspecific genetic diversity was likely caused by a direct impact of one facet of biodiversity on the other. We speculated that genetic drift led to morphological, physiological or behavioural divergences among populations thereby influencing the structure of the local ecosystems and communities (Vellend and Geber 2005).

In Chapter III, we found no spatial covariation between intraspecific genetic and phenotypic α-diversity. The neutral processes driving neutral genetic diversity (e.g. genetic drift) and the adaptive processes driving phenotypic diversity (i.e. local adaptation) were

expected to spatially covary within riverine network (because of the gradients in habitat sizes and environmental conditions notably, Vannote et al. 1980). We found support for this hypothesis: in *G. occitaniae*, genetic α-diversity and phenotypic α-diversity were both found to be lower in upstream sites, although these patterns probably originate from different processes. However, no correlation emerged from this covariation, therefore suggesting that the influence of other processes (related to connectivity) are strong enough to break spatial covariation between these two facets of diversity in this species.

At the β-level, we found a positive correlation between genetic diversity and phenotypic diversity in *G. occitaniae* but not in *P. phoxinus*. This correlation appeared to originate from a direct impact of one facet of intraspecific diversity on the other, and we hypothesized that positive assortative mating (i.e. the propensity to mate with phenotypically similar individuals), that has been shown to be particularly strong in fish (Jiang et al. 2013), could explain this direct relation between phenotypic and genetic differentiation (Wang and Summers 2010).

Studying multiple facets of biodiversity within integrative frameworks has allowed us to uncover complex and often unexpected processes driving biodiversity patterns. While introducing the novel framework of Species-Genetic Diversity Correlations, Vellend (2005) stated that the tradition of treating interspecific and intraspecific diversity as independent phenomena in community ecology and population genetics was irrelevant and outdated. Similarly, we advocate for a greater integration across disciplinary boundaries in order to "get the big picture" and enhance our understanding of the causes and consequences of biodiversity, with the eventual objective of proposing to decision-makers more efficient and sustainable methods to limit biodiversity loss.

The use of causal modeling in ecology and evolution studies

All the previous results were obtained through causal modeling methods. Without these methods, we would have been incapable of disentangling the direct and indirect paths linking environmental variables and diversity descriptors. The assets of causal modeling for testing ecological and evolutionary hypotheses have long been acknowledged (Mitchell 1992; Grace et al. 2010). However, despite its relevance for analyzing complex observational data,

causal modeling is still only occasionally used in ecology.

Until now, the application of causal modeling methods were restricted to point-summary data (one value per sampling site) (but see Cushman et al. 2013). However, when assessing the impact of landscape features such as topographic distance on diversity (Manel et al. 2003), the statistical analysis of relations among pairwise matrices is necessary and the use of causal modeling was then strongly restricted (Cushman et al. 2013). In Chapter I, we developed new approaches allowing the application of two widely used causal modeling methods (namely maximum-likelihood-based path analysis and the d-sep test) to data in the form of pairwise matrices.

As goes with all statistical methods, causal modeling must be used with caution. Notably, causal modeling relies solely on the causal hypotheses depicted within the initial causal model and cannot provide information beyond the stated hypotheses. As a consequence, inferred relationships among variables must be considered with caution because some important but unknown (or unmeasured) variables may have been overlooked. Additionally, causal modeling cannot be confidently used as a data mining procedure, as implicit *a priori* hypotheses are fundamental to avoid spurious conclusions (Shipley 2009; Legendre and Legendre 2012; Prunier et al. 2015). Keeping these limitations in mind, we believe that causal modeling is a powerful explanatory tool and we hope that our contribution will favour its use in future ecology and evolution studies.

References

Adams, R. I., and E. A. Hadly. 2013. Genetic diversity within vertebrate species is greater at lower latitudes. Evolutionary Ecology 27:133–143.

Altermatt, F. 2013. Diversity in riverine metacommunities: a network perspective. Aquatic Ecology 47:365–377.

Antonovics, J. 1976. The Input from Population Genetics: "The New Ecological Genetics." Systematic Botany 1:233–245.

Badyaev, A. V. 2005. Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. Proceedings of the Royal Society of London B: Biological Sciences 272:877–886.

Baselga, A. 2010. Multiplicative partition of true diversity yields independent alpha and beta components; additive partition does not. Ecology 91:1974–1981.

Begon, M., C. R. Townsend, and J. L. Harper. 2006. Ecology: from individuals to ecosystems (4th edition.). Wiley-Blackwell, Malden, MA.

Blanchet, S., M. R. Helmus, S. Brosse, and G. Grenouillet. 2014. Regional vs local drivers of phylogenetic and species diversity in stream fish communities. Freshwater Biology 59:450–462.

Blanchet, S., J. G. Prunier, and H. De Kort. 2017. Time to Go Bigger: Emerging Patterns in Macrogenetics. Trends in Genetics 33:579–580.

Bolnick, D. I., P. Amarasekare, M. S. Araújo, R. Bürger, J. M. Levine, M. Novak, V. H. W. Rudolf, et al. 2011. Why intraspecific trait variation matters in community ecology. Trends in Ecology & Evolution 26:183–192.

Bolnick, D. I., R. Svanbäck, J. A. Fordyce, L. H. Yang, J. M. Davis, C. D. Hulsey, and M. L. Forister. 2003. The ecology of individuals: incidence and implications of individual specialization. The American Naturalist 161:1–28.

Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, et al. 2010. Global Biodiversity: Indicators of Recent Declines. Science 328:1164–1168.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, et al. 2012. Biodiversity loss and its impact on humanity. Nature 486:59–67.

Carrara, F., F. Altermatt, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. Dendritic connectivity controls biodiversity patterns in experimental metacommunities. Proceedings of the National

Academy of Sciences 109:5761–5766.

Carrara, F., A. Rinaldo, A. Giometto, and F. Altermatt. 2014. Complex Interaction of Dendritic Connectivity and Hierarchical Patch Size on Biodiversity in River-Like Landscapes. The American Naturalist 183:13–25.

Chave, J. 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? Ecology Letters 16:4–16.

Connor, E. F., and E. D. McCoy. 1979. The Statistics and Biology of the Species-Area Relationship. The American Naturalist 113:791–833.

Cushman, S. A., T. N. Wasserman, E. L. Landguth, and A. J. Shirk. 2013. Re-evaluating causal modeling with mantel tests in landscape genetics. Diversity 5:51–72.

Díaz, S., J. Fargione, F. S. C. Iii, and D. Tilman. 2006. Biodiversity Loss Threatens Human Well-Being. PLOS Biology 4:e277.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, et al. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81:163.

Eduardo, A. A. 2016. Multiple dimensions of biodiversity and ecosystem processes: Exploring the joint influence of intraspecific, specific and interspecific diversity. Journal of Theoretical Biology 404:215–221.

Fagan, W. F. 2002. Connectivity, Fragmentation, and Extinction Risk in Dendritic Metapopulations. Ecology 83:3243–3249.

Frankham, R. 1996. Relationship of Genetic Variation to Population Size in Wildlife. Conservation Biology 10:1500–1508.

Freestone, A. L., R. W. Osman, G. M. Ruiz, and M. E. Torchin. 2011. Stronger predation in the tropics shapes species richness patterns in marine communities. Ecology 92:983–993.

Fu, C., J. Wu, X. Wang, G. Lei, and J. Chen. 2004. Patterns of diversity, altitudinal range and body size among freshwater fishes in the Yangtze River basin, China. Global Ecology and Biogeography 13:543–552.

Gaston, K. J. 2000. Global patterns in biodiversity. Nature 405:220–227.

Ghalambor, C. K., J. K. McKAY, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394–407.

Gotelli, N. J., M. J. Anderson, H. T. Arita, A. Chao, R. K. Colwell, S. R. Connolly, D. J. Currie, et al. 2009. Patterns and causes of species richness: a general simulation model for

macroecology. Ecology Letters 12:873–886.

Grace, J. B. 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press, Cambridge.

Grace, J. B., T. M. Anderson, H. Olff, and S. M. Scheiner. 2010. On the specification of structural equation models for ecological systems. Ecological Monographs 80:67–87.

Grant, E. H. C., W. H. Lowe, and W. F. Fagan. 2007. Living in the branches: population dynamics and ecological processes in dendritic networks. Ecology Letters 10:165–175.

Grant, E. H. C., J. D. Nichols, W. H. Lowe, and W. F. Fagan. 2010. Use of multiple dispersal pathways facilitates amphibian persistence in stream networks. Proceedings of the National Academy of Sciences 107:6936–6940.

Graves, T. A., P. Beier, and J. A. Royle. 2013. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. Molecular Ecology 22:3888–3903.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Hanski, I., T. Mononen, and O. Ovaskainen. 2011. Eco-Evolutionary Metapopulation Dynamics and the Spatial Scale of Adaptation. The American Naturalist 177:29–43.

Heino, J. 2002. Concordance of species richness patterns among multiple freshwater taxa: a regional perspective. Biodiversity & Conservation 11:137–147.

Heino, J., R. Virkkala, and H. Toivonen. 2009. Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. Biological Reviews 84:39–54.

Hillebrand, H. 2004. On the Generality of the Latitudinal Diversity Gradient. The American Naturalist 163:192–211.

Hoffman, J. I., F. Simpson, P. David, J. M. Rijks, T. Kuiken, M. A. S. Thorne, R. C. Lacy, et al. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences 111:3775–3780.

Hooper, D. U., E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A. Hungate, K. L. Matulich, A. Gonzalez, et al. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486:105–108.

Hopper, G. W., R. L. Morehouse, and M. Tobler. 2015. Body shape variation in two species of darters ([i]Etheostoma[/i], Percidae) and its relation to the environment. Ecology of Freshwater Fish 26:4–18.

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecology Letters 11:609–623.

Hutchison, D., and A. Templeton. 1999. Correlation of pairwise genetic and geographic distance measure: inferring the relative influences of gene flow and drift on distribution of genetic variability. Evolution 53:1898–1914.

Jackson, D. A., P. R. Peres-Neto, and J. D. Olden. 2001. What controls who is where in freshwater fish communities - the roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences 58:157–170.

Jiang, Y., D. I. Bolnick, and M. Kirkpatrick. 2013. Assortative Mating in Animals. The American Naturalist 181:E125–E138.

Jung, V., C. H. Albert, C. Violle, G. Kunstler, G. Loucougaray, and T. Spiegelberger. 2013. Intraspecific trait variability mediates the response of subalpine grassland communities to extreme drought events. Journal of Ecology 102:45–53.

Kawecki, T. J., and R. D. Holt. 2002. Evolutionary consequences of asymmetric dispersal rates. The American Naturalist 160:333–347.

Lande, R. 1996. Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. Oikos 76:5–13.

Lawton, J. H. 1996. Patterns in Ecology. Oikos 75:145–147.

Legendre, P., and L. F. J. Legendre. 2012. Numerical Ecology. Elsevier, Amsterdam.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17:183–189.

Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. Ecology 73:1943–1967.

Loreau, M. 2000*a*. Biodiversity and ecosystem functioning: recent theoretical advances. Oikos 91:3–17.

———. 2000*b*. Are communities saturated? On the relationship between α, β and γ diversity. Ecology Letters 3:73–76.

———. 2010. Linking biodiversity and ecosystems: towards a unifying ecological theory. Philosophical Transactions of the Royal Society B: Biological Sciences 365:49–60.

Lowe, W. H., R. P. Kovach, and F. W. Allendorf. 2017. Population Genetics and Demography Unite Ecology and Evolution. Trends in Ecology & Evolution 32:141–152.

Maclaurin, J., and K. Sterelny. 2008. What is biodiversity? University of Chicago Press, Chicago, IL.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189–197.

Mimura, M., T. Yahara, D. P. Faith, E. Vázquez-Domínguez, R. I. Colautti, H. Araki, F. Javadi, et al. 2017. Understanding and monitoring the consequences of human impacts on intraspecific variation. Evolutionary Applications 10:121–139.

Mitchell, R. J. 1992. Testing Evolutionary and Ecological Hypotheses Using Path Analysis and Structural Equation Modelling. Functional Ecology 6:123–129.

Mittelbach, G. G., D. W. Schemske, H. V. Cornell, A. P. Allen, J. M. Brown, M. B. Bush, S. P. Harrison, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecology Letters 10:315–331.

Moran, E. V., F. Hartig, and D. M. Bell. 2015. Intraspecific trait variation across scales: implications for understanding global change responses. Global Change Biology 22:137–150.

Morrissey, M. B., and D. T. de Kerckhove. 2009. The Maintenance of Genetic Variation Due to Asymmetric Gene Flow in Dendritic Metapopulations. The American Naturalist 174:875–889.

Muneepeerakul, R., E. Bertuzzo, A. Rinaldo, and I. Rodriguez-Iturbe. 2008. Patterns of vegetation biodiversity: The roles of dispersal directionality and river network structure. Journal of Theoretical Biology 252:221–229.

Odling-Smee, F. J., K. N. Laland, and M. W. Feldman. 2003. Niche Construction: The Neglected Process in Evolution. Princeton University Press, Princeton, NJ and Oxford.

Paz-Vinas, I., and S. Blanchet. 2015. Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. Journal of Evolutionary Biology 28:986–994.

Paz-Vinas, I., G. Loot, V. M. Stevens, and S. Blanchet. 2015. Evolutionary processes driving spatial patterns of intra-specific genetic diversity in river ecosystems. Molecular Ecology 24:4586–4604.

Paz-Vinas, I., E. Quéméré, L. Chikhi, G. Loot, and S. Blanchet. 2013. The demographic history of populations experiencing asymmetric gene flow: combining simulated and empirical data. Molecular Ecology 22:3279–3291.

Pianka, E. R. 1966. Latitudinal Gradients in Species Diversity: A Review of Concepts. The American Naturalist 100:33–46.

Prunier, J. G., M. Colyn, X. Legendre, K. F. Nimon, and M. C. Flamand. 2015. Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. Molecular Ecology 24:263–283.

Prunier, J. G., V. Dubut, L. Chikhi, and S. Blanchet. 2017. Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. Methods in Ecology and Evolution Early view.

Rahbek, C. 1995. The elevational gradient of species richness: a uniform pattern? Ecography 18:200–205.

Rey, O., E. Danchin, M. Mirouze, C. Loot, and S. Blanchet. 2016. Adaptation to Global Change: A Transposable Element–Epigenetics Perspective. Trends in Ecology & Evolution 31:514–526.

Reyjol, Y., B. Hugueny, D. Pont, P. G. Bianco, U. Beier, N. Caiola, F. Casals, et al. 2007. Patterns in species richness and endemism of European freshwater fish. Global Ecology and Biogeography 16:65–75.

Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters 7:1–15.

Ricklefs, R. E., and G. L. Miller. 2000. Ecology (4th edition.). W.H. Freeman & Co, New York, NY.

Ridley, M. 2003. Evolution (3rd edition.). Blackwell Scientific Publishing, Malden, MA.

Ronce, O., and M. Kirkpatrick. 2001. When Sources Become Sinks: Migrational Meltdown in Heterogeneous Habitats. Evolution 55:1520–1531.

Schlosser, I. J. 1990. Environmental variation, life history attributes, and community structure in stream fishes: Implications for environmental management and assessment. Environmental Management 14:621–628.

Shipley, B. 2000*a*. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge.

———. 2000*b*. A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling 7:206–218.

———. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90:363–368.

Strayer, D. L., and D. Dudgeon. 2010. Freshwater biodiversity conservation: recent progress and future challenges. Journal of the North American Benthological Society 29:344–358.

Taberlet, P., N. E. Zimmermann, T. Englisch, A. Tribsch, R. Holderegger, N. Alvarez, H. Niklfeld, et al. 2012. Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. (M. Vellend, ed.)Ecology Letters 15:1439–1448.

Vander Vorste, R., P. McElmurray, S. Bell, K. Eliason, and B. Brown. 2017. Does Stream Size

Really Explain Biodiversity Patterns in Lotic Systems? A Call for Mechanistic Explanations. Diversity 9:26.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130–137.

Vellend, M. 2003. Island Biogeography of Genes and Species. The American Naturalist 162:358–365.

———. 2005. Species diversity and genetic diversity: parallel processes and correlated patterns. The American Naturalist 166:199–215.

Vellend, M., and M. A. Geber. 2005. Connections between species diversity and genetic diversity. Ecology Letters 8:767–781.

Vellend, M., G. Lajoie, A. Bourret, C. Múrria, S. W. Kembel, and D. Garant. 2014. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. Molecular Ecology 23:2890–2901.

Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, et al. 2010. Global threats to human water security and river biodiversity. Nature 467:555–561.

Wang, I. J., and K. Summers. 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Molecular Ecology 19:447–458.

Whittaker, R. H. 1972. Evolution and Measurement of Species Diversity. Taxon 21:213–251.

Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28:453–470.

Williams, P., M. Whitfield, J. Biggs, S. Bray, G. Fox, P. Nicolet, and D. Sear. 2004. Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. Biological Conservation 115:329–341.

Willig, M. R., D. M. Kaufman, and R. D. Stevens. 2003. Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. Annual Review of Ecology, Evolution, and Systematics 34:273–309.

Wright, S. 1934. The Method of Path Coefficients. The Annals of Mathematical Statistics 5:161–215.

AUTHOR:
LISA FOURTUNE

TITLE:
Inter- and intra-specific diversity patterns in dendritic river networks

THESIS SUPERVISOR:
SIMON BLANCHET

DATE AND LOCATION OF THE DEFENCE:
January, 12th 2018 at Station d'Écologie Théorique et Expérimentale de Moulis, France

---

SUMMARY:
The aim of this thesis was to characterized the spatial patterns of inter- and intraspecific diversity within riverine networks, to better understand the ecological and evolutionary processes underlying them and to explore how the different facets of biodiversity interact with one another. First, I developed novel statistical approaches allowing the application of causal modeling to data in the form of pairwise matrices, thus allowing the study within integrative frameworks of several biodiversity facets at the alpha and beta levels. I then studied integratively the patterns of interspecific and intraspecific genetic diversity and of intraspecific genetic and intraspecific phenotypic diversity within the Garonne-Dordogne river basin. Finally, I used an eco-evolutionary metapopulation dynamics model to assess the impacts of the structure and environmental gradients that characterize riverine networks on local adaptation.

KEYWORDS:
Diversity patterns, dendritic river networks, causal modeling,
interspecific diversity, intraspecific diversity

DOCTORAL SCHOOL AND DISCIPLINE:
ED SEVAB - Ecology and Evolutionnary Biology

RESEARCH UNIT:
UMR 5321 - Station d'Écologie Théorique et Expérimentale
2, route du CNRS
09200 Moulis, France

AUTEUR :
LISA FOURTUNE

TITRE :
Patrons de diversité inter- et intraspécifique dans les réseaux dendritiques d'eau douce : implications pour leur fonctionnement et leur conservation

DIRECTEUR DE THÈSE :
SIMON BLANCHET

DATE ET LIEU DE SOUTENANCE :
Le 12 janvier 2018 à la Station d'Écologie Théorique et Expérimentale de Moulis, France

---

RÉSUMÉ :

L'objectif de cette thèse a été de caractériser les patrons spatiaux de diversité inter- et intraspécifique au sein des réseaux dendritiques, d'expliciter les processus évolutifs et écologiques qui les sous-tendent, et d'isoler les possibles covariations spatiales et interactions existant entre ces différentes facettes de biodiversité. Pour cela, j'ai tout d'abord développé de nouvelles méthodes statistiques permettant l'analyse, par des modèles causaux, de données sous la forme de matrices de distances, afin de pouvoir analyser plusieurs facettes de biodiversité dans un unique cadre statistique au niveau alpha et bêta. J'ai par la suite étudié de manière intégrative les patrons de diversité interspécifique et intraspécifique génétique d'une part, et intraspécifique génétique et intraspécifique phénotypique d'autre part, au sein du bassin versant Garonne-Dordogne. Enfin, j'ai utilisé un modèle de dynamique éco-évolutive afin d'étudier l'impact de la structure et des gradients environnementaux caractérisant les réseaux dendritiques sur l'adaptation locale au sein de ces réseaux.

MOTS-CLÉS :
Patrons spatiaux de diversité, réseaux dendritiques, analyses causales, diversité interspécifique, diversité intraspécifique

ÉCOLE DOCTORALE ET DISCIPLINE :
ED SEVAB - Écologie et Évolution

UNITÉ DE RECHERCHE :
UMR 5321 - Station d'Écologie Théorique et Expérimentale
2, route du CNRS
09200 Moulis, France